

# 计算机理论基础选讲

谢浩哲

[cshzxie@gmail.com](mailto:cshzxie@gmail.com)

# 计算机软件

# 计算机软件概述

- ▶ 软件(Software)是一系列按照特定顺序组织的电脑数据和指令的集合. **一般来讲软件被划分为系统软件、支撑软件和应用软件.** 其中系统软件为计算机使用提供最基本的功能, 但是并不针对某一特定应用领域. 而应用软件则恰好相反, 不同的应用软件根据用户和所服务的领域提供不同的功能.
- ▶ 软件并不只是包括可以在计算机上运行的应用程序, 与这些应用程序相关的文档和数据, 一般也被认为是软件的一部分.
- ▶ **Software = Program + Documents + Data**

# 计算机软件概述

## 应用软件

- Office, WPS
- MySQL, Oracle, SQL Server
- Photoshop, Illustrator
- IE, Chrome, Firefox, Opera

## 支撑软件

- 支撑软件是支撑软件开发、维护和运行的软件.

## 系统软件

- 系统软件是负责管理计算机系统中各种独立的硬件, 使得它们可以协调工作.

# 支撑软件概述

## 外壳交互程序

- Shell

## 软件开发工具

- Free Pascal
- Dev C++
- Eclipse

## 文本编辑器

- Vim
- Emacs

## 资源管理工具

- Explorer

# 操作系统概述

- ▶ 操作系统(Operating System)
  - ▶ 操作系统是是一管理电脑硬件与软件资源的程序,同时也是计算机系统的内核与基石. 操作系统是一个庞大的管理控制程序, **大致包括4个方面的管理功能: 进程管理、存储管理、设备管理、文件管理**. 以现代观点而言, 标准个人电脑操作系统应提供以下功能: 进程管理(Processing Management); 存储空间管理(Memory Management); 文件系统(File System); 网络通讯(Networking); 安全机制(Security); 使用者界面(User Interface); 驱动程序(Device Drivers).

# 操作系统架构概述

Application

Shell

GUI

System Call

Kernel

Drivers

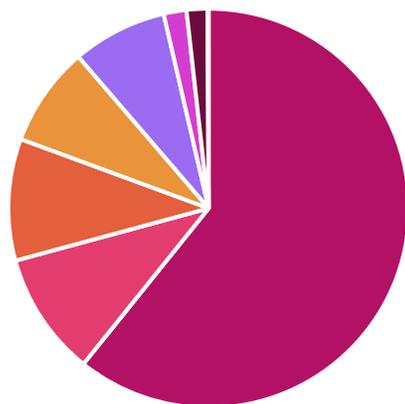
Hardware

# 操作系统分类

- ▶ Categorized based on the types of computers they control and the sort of applications they support.
  - ▶ Real-time operating system
  - ▶ Single-user, single task
  - ▶ Single-user, multi-tasking
  - ▶ Multi-user
- ▶ Reference: <http://computer.howstuffworks.com/operating-system3.htm>

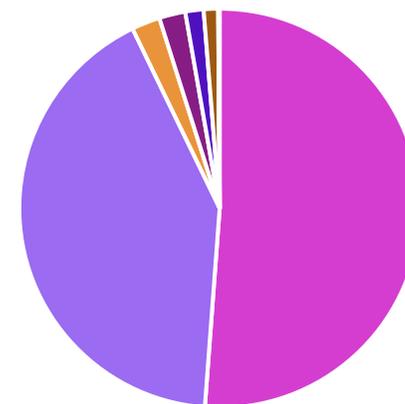
# 典型的操作系统

Desktop OS(July, 2015)



- Windows 7
- Windows 10
- Windows 8.x
- Windows XP
- Mac OS X
- Windows Vista
- Linux
- Other

Mobile OS(July, 2015)



- Android
- iOS
- Windows Phone
- Java ME
- Sybian
- BlackBerry
- Kindle
- Other

Reference: [https://en.wikipedia.org/wiki/Usage\\_share\\_of\\_operating\\_systems](https://en.wikipedia.org/wiki/Usage_share_of_operating_systems)

# 典型的操作系统

## ▶ Desktop Operating System

### ▶ Windows Family

- ▶ Latest OS: Windows 10

- ▶ History Version: Windows 8.x, Windows 7, Windows Vista, Windows XP, Windows 2000, Windows ME, Windows 9x, Windows 3.x, ..., DOS

### ▶ Windows Server Family

- ▶ Latest OS: Windows Server 2012 R2

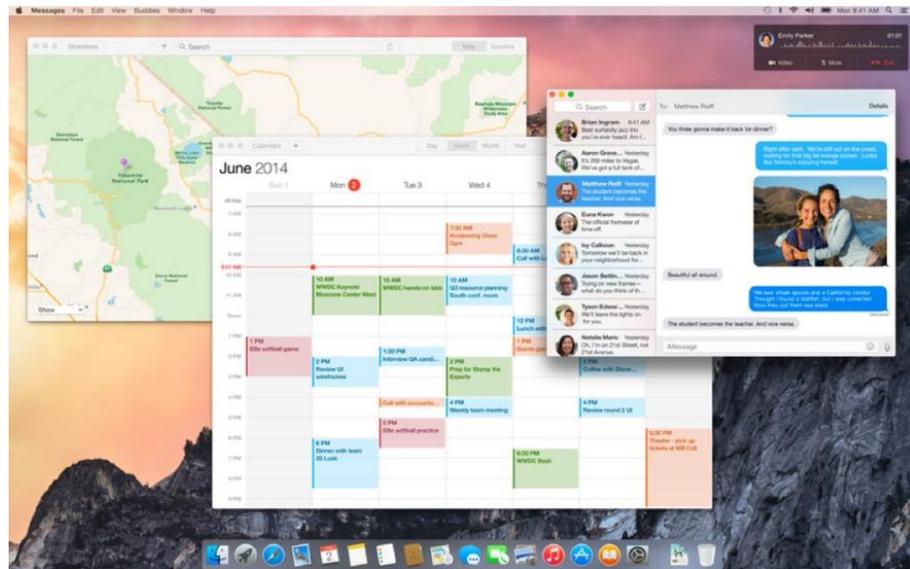
- ▶ History Version: Windows Server 2008 (R2), Windows Home Server, Windows Server 2003, Windows 2000 Server



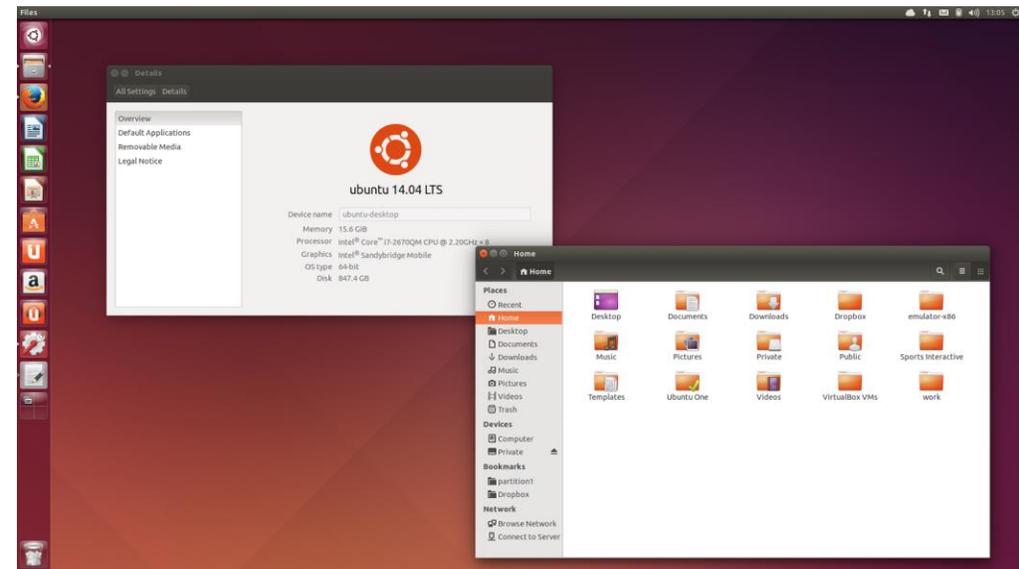
# 典型的操作系统

- ▶ Desktop Operating System
  - ▶ Unix Family
    - ▶ Mac OS
    - ▶ IBM AIX
    - ▶ Solaris
  - ▶ Linux Family
    - ▶ Debian Series: Debian (8), Ubuntu (15.04), Linux Mint (18), etc.
    - ▶ Red Hat Series: Red Hat (7), Fedora (22), Cent OS (7), etc.
    - ▶ etc.

# Mac OS + Ubuntu Screenshot



Mac OS 10.10



Ubuntu 14.04 LTS

# Shell Commands

## ▶ DOS Command

- ▶ cd
- ▶ dir
- ▶ mkdir(md)
- ▶ rmdir(rd)
- ▶ del

## ▶ Unix/Linux Shell Command

- ▶ cd
- ▶ ls
- ▶ mkdir
- ▶ rmdir
- ▶ rm
- ▶ Feature in Shell: Pipe
  - ▶ eg: netstat -anp | grep 80

# 典型的操作系统

## ▶ Mobile Operating System

### ▶ iOS

- ▶ Latest Version: 8.4
- ▶ Develop Language: Objective-C/ Swift

### ▶ Android

- ▶ Latest Version: 5.1
- ▶ Develop Language: Java SE with Android SDK

### ▶ Windows Phone (Windows RT)

- ▶ Latest Version: 10
- ▶ Develop Language: C#, VB.NET, etc.

# 文件系统

- ▶ 文件系统是一种用于向用户提供底层数据存取机制。它将设备中的空间划分为特定大小的块(扇区), 一般每块512字节。数据存储在那些块中, 大小被修正为占用整数个块。由文件系统软件来负责将这些块组织为文件和目录, 并记录哪些块被分配给了哪个文件, 以及哪些块没有被使用。
- ▶ 常见的文件系统
  - ▶ FAT、exFAT、NTFS、HFS、HFS+、ext2、ext3、ext4, etc.

# 文件类型

- ▶ 文件类型是指计算机为了存储信息而使用的对信息的特殊编码方式. 每一种文件格式通常会有一种或多种扩展名可以用来识别, 但也可能没有扩展名.
- ▶ 从程序的角度来看, 文件是数据流, 文件系统为每一种文件格式规定了访问的方法. 不同的操作系统都习惯性的采用各自的方式解决这个问题.

# 识别文件类型

## ▶ 扩展名

- ▶ 扩展名是指文件名中, 最后一个点(.)号后的字母序列. 例如, HTML文件通过.htm或.html扩展名识别.

## ▶ 特征签名

- ▶ 一种广泛应用在UNIX及其派生的操作系统上的方法是将一个特殊的数字存放在文件的特定位置里. 最初这个数字一般是文件开始处的2个字节. 现在一般是将任何可以独一无二字符序列都可以作为特征签名.

# 常见的文件扩展名

- ▶ Office文档
  - ▶ doc(x), xls(x), ppt(x), mdb, etc.
- ▶ 音频文件
  - ▶ wma, wav, mp3, aac, rm, etc.
- ▶ 视频文件
  - ▶ wmv, asf, mp4, avi, mpg, mov, rmvb, flv, f4v, etc.
- ▶ 图像文件
  - ▶ jpg, bmp, png, gif, tif, tiff, etc.
- ▶ 压缩文件
  - ▶ rar, zip, 7z, tar, gz, bz, xz, etc.

# 数据库系统

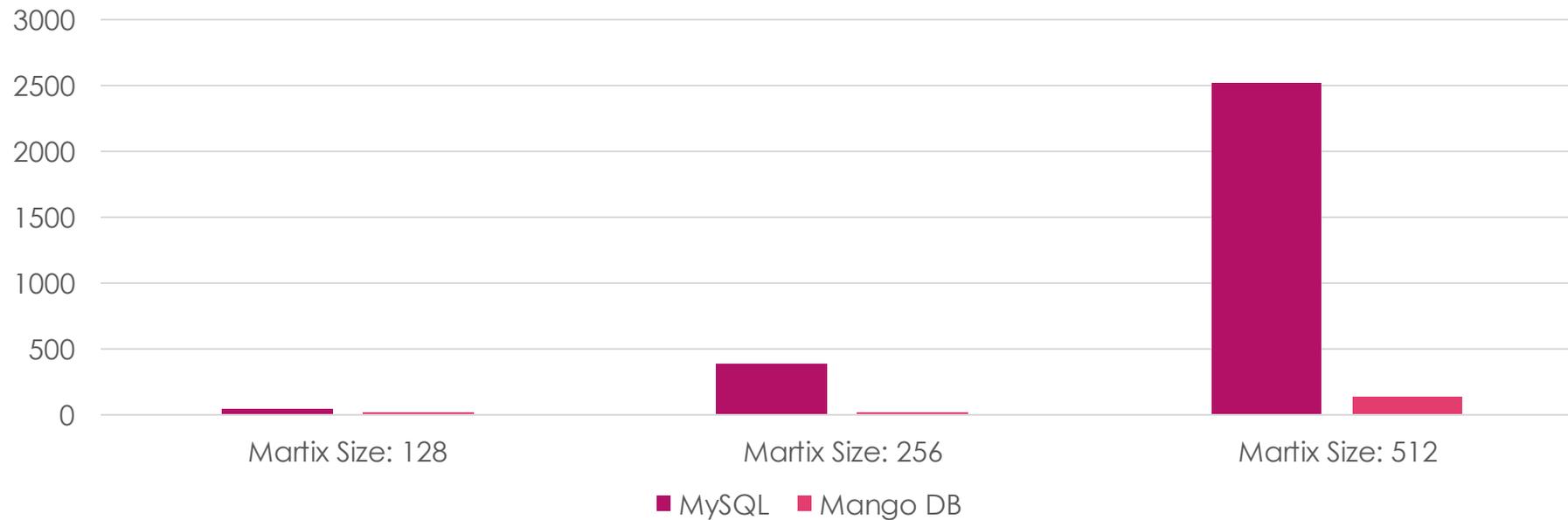
- ▶ 数据库, 简单来说本身可视为电子化的文件柜——存储电子文件的处所, 用户可以对文件中的数据运行**新增、截取、更新、删除**等操作.
- ▶ 数据库指的是以一定方式储存在一起、能为多个用户共享、具有尽可能小的冗余度、与应用程序彼此独立的数据集合.
- ▶  $DBS = DB + DBMS + DBA + Application + User$

# 数据库管理系统

- ▶ A "database management system" (DBMS) is a suite of computer software providing the interface between users and a database or databases.
- ▶ **Example:**
  - ▶ **Relational DBMS:** MySQL(Maria DB), Oracle, SQL Server, DB2, Sybase, FoxPro, Access, etc.
  - ▶ **NoSQL:** Mango DB, Cassandra, Redis, Hive, Hbase, etc.

# Why NoSQL?

Query Time (Smaller is better)



# NoSQL is Better?

## ▶ CAP 原则

▶ Consistency

▶ Availability

▶ Partition tolerance

# 编程语言

- ▶ Computer programs, known as software, are instructions to the computer.
- ▶ You tell a computer what to do through programs. Without programs, a computer is an empty machine. Computers do not understand human languages, so you need to use computer languages to communicate with them.
- ▶ Programs are written using programming languages.

# 编程语言分类

- ▶ 按语言级别分类

- ▶ 低级语言

- ▶ 机器语言(Machine Language)

- ▶ 汇编语言(Assembly Language)

- ▶ 高级语言

- ▶ 面向过程的语言

- ▶ 面向对象的语言

# 编程语言分类

## ▶ 高级语言

### ▶ 面向过程语言

- ▶ 基本单元: 函数/过程

- ▶ Example: FORTRAN、Basic、C、Pascal

### ▶ 面向对象语言

- ▶ 基本单元: 类(class)

- ▶ 面向对象的三个特性：封装(Encapsulation)、继承(Inheritance)和多态(Polymorphism)

- ▶ Example: **Simula**, SmallTalk, C++, Java, C#, Visual Basic.NET, Objective-C, PHP, Python, Ruby, JavaScript, Objective-Pascal

# 编程语言分类

## ▶ 按编译方式分类

### ▶ 解释型语言

▶ Example: Basic, PHP, Python, Ruby, JavaScript, Bash (Shell), R

### ▶ 编译型语言

▶ Example: C, C++, Pascal

### ▶ 解释 + 编译型语言

▶ Example: Java, C#, VB.NET

# 计算机硬件

# 单位换算

- ▶ 1 Byte = 8 Bit
  - ▶ 1 KB = 1024 Byte
  - ▶ 1 MB = 1024 KB
  - ▶ 1 GB = 1024 MB
- |      |   |         |
|------|---|---------|
| 1TB  | = | 1024 GB |
| 1 PB | = | 1024 TB |
| 1 EB | = | 1024 PB |

# 阿兰·麦席森·图灵

## ▶ 阿兰·麦席森·图灵 (Alan Mathison Turing)

- ▶ 6月23日生于英国伦敦. 是英国著名的数学家和逻辑学家, 被称为计算机科学之父、人工智能之父, 是计算机逻辑的奠基者, 提出了“图灵机”和“图灵测试”等重要概念. 人们为纪念其在计算机领域的卓越贡献而设立“图灵奖”.

## ▶ 图灵机

- ▶ 只要按一定步骤对数字进行处理、传输和存储三种操作, 就能在理论上解决任何数学计算问题.
- ▶ 1936年, 主要研究可计算理论, 并提出“图灵机”的构想.



# 约翰·冯·诺依曼

- ▶ 约翰·冯·诺依曼 (John Von Neumann)
  - ▶ 1903年12月28日, 在布达佩斯诞生了一位神童, 这不仅给这个家庭带来了巨大的喜悦, 也值得整个计算机界去纪念. 正是他, 开创了现代计算机理论, 其体系结构沿用至今, 而且他早在40年代就已预见到计算机建模和仿真技术对当代计算机将产生的意义深远的影响.
  - ▶ 1945年6月, 冯·诺伊曼等人, 联名发表了一篇长达101页纸的报告, 即计算机史上著名的“101页报告”, 明确规定用二进制替代十进制运算.
  - ▶ 将计算机分成五大组件, 这一卓越的思想为电子计算机的逻辑结构设计奠定了基础, 已成为计算机设计的基本原则. 1951年, EDVAC计算机宣告完成. 由于他在计算机逻辑结构设计上的伟大贡献, 他被誉为“计算机之父”.



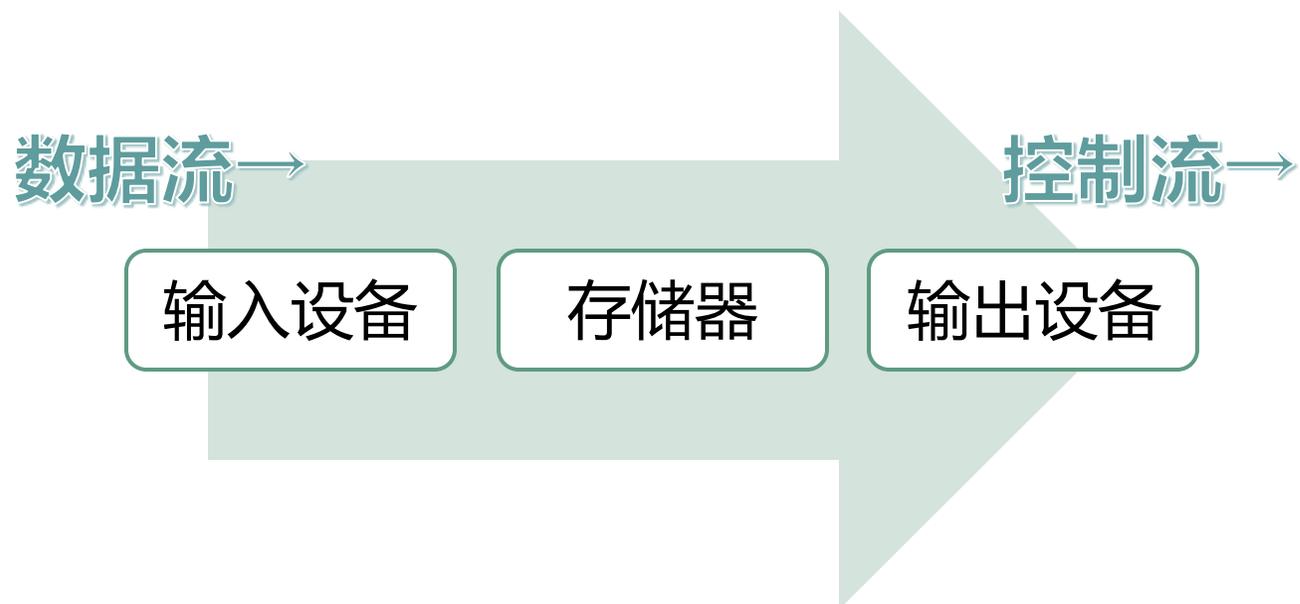
# 计算机发展史

## ▶ 第一台电子计算机

- ▶ 世界上第一台现代电子计算机“埃尼阿克”(ENIAC), 诞生于1946年2月14日的美国宾夕法尼亚大学, 并于次日正式对外公布.
- ▶ 第二次世界大战期间, 美国军方要求宾州大学Mauchly博士和他的学生爱克特(Eckert) 设计以真空管取代继电器的"电子化"电脑ENIAC, 电子数字积分器与计算器, 目的是用来计算炮弹弹道. 这部机器使用了18800个真空管, 长50英尺, 宽30英尺, 占地1500平方英尺, 重达30吨. 它的计算速度快, 每秒可从事5000次的加法运算, 运作了九年之久. 据传ENIAC每次一开机, 整个费城西区的电灯都为之黯然失色.

# 计算机硬件概述

- ▶ 计算机硬件由五大部分组成:
  - ▶ 运算器、控制器、存储器、输入设备、输出设备.



# 计算机的特点与分类

## ▶ 计算机的特点:

- ▶ 运算速度快、计算精度高、存储能力强、具有逻辑判断能力、具有自动执行能力

## ▶ 计算机的分类

- ▶ 巨型机, 大型机, 中型机, 小型机, 工作站(WorkStation), 微型机(PC), etc.

# 主板

- ▶ 主板是硬件的主要部分. 它是一况大的方形板, 上面有复杂的电路连接到电脑的其他部分, 内含:
  - ▶ 中央处理器(CPU): 处理电脑内大部分的计算, 其被认为是电脑的大脑. 新型的中央处理器包含图形处理器.
  - ▶ 芯片组: 作为中央处理器跟系统内其他组件两者间的联络工具.
  - ▶ 内存(RAM): 存储正在运行的操作系统以及所有正在运行的程序.
  - ▶ BIOS或EFI: 包含固件及电源管理系统
  - ▶ 内部总线: 用来连接中央处理器和电脑内部其他组件或是显卡, 声卡.



# CPU(中央处理器)

- ▶ CPU, (Central Processing Unit), 电子计算机的主要设备之一. 其功能主要是解释计算机指令以及处理计算机软件中的数据.
- ▶ 冯·诺伊曼CPU的运作原理可分为四个阶段:
  - ▶ 提取、解码、执行和写回.



# Intel

## ▶ Intel CPU主要产品:

### ▶ 桌面型

- ▶ 8086、8088、80286、80386、80486、Pentium、Pentium Pro、Pentium MMX、Pentium II、Celeron、Pentium III、Pentium 4、Pentium D、Celeron D、Core、Core 2、Pentium Dual-Core、Celeron Dual-Core、Core i3、Core i5、Core i7

### ▶ 笔记本电脑

- ▶ Pentium III Mobile、Pentium 4 Mobile、Pentium M、Celeron M、Core、Core 2、Atom、Core i3、Core i5、Core i7

### ▶ 服务器

- ▶ Xeon、Itanium



# AMD

## ▶ AMD CPU主要产品:

- ▶ Geode
- ▶ Sempron
- ▶ Athlon 64
- ▶ Athlon X2
- ▶ Turion 64
- Turion 64 X2
- Athlon 64 FX
- Phenom
- Phenom II
- Opteron



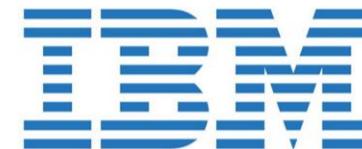
# IBM POWER

## ▶ IBM POWER概述

- ▶ IBM POWER是RISC处理器架构的一种,由IBM设计,全称为“Performance Optimization With Enhanced RISC”。POWER系列微处理器在不少IBM服务器、超级计算机、小型机及工作站中。

## ▶ IBM POWER主要产品:

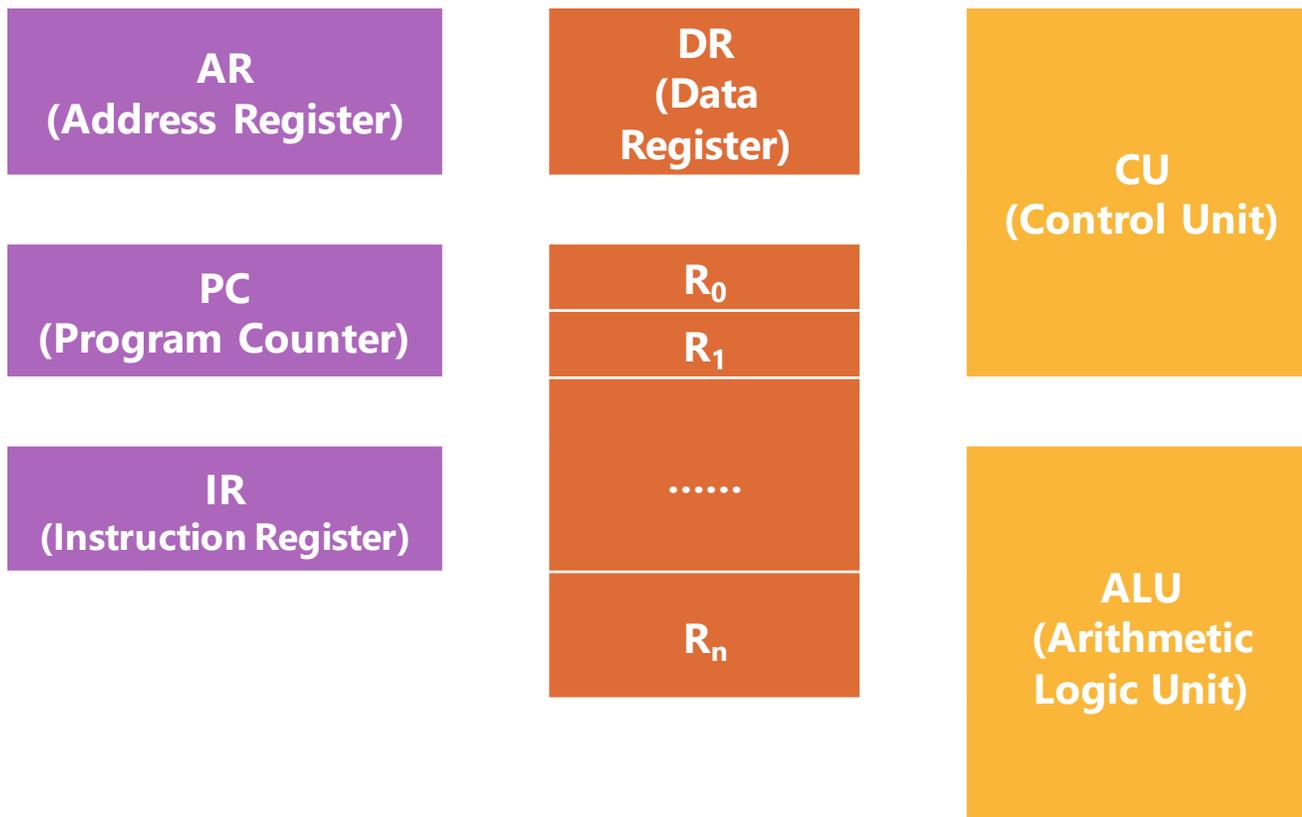
- ▶ POWER 1, POWER 2, ..., POWER 7



# CPU基本结构

- ▶ CPU是由算术逻辑单元、控制器和寄存器组成
  - ▶ CPU = ALU + CU + Register (+ Cache)
- ▶ 运算器
  - ▶ 功能:算术运算、逻辑运算
  - ▶ 部件:算术逻辑单元、累加器、路径转换器、数据总线
- ▶ 控制器
  - ▶ 功能:复位(Reset)、使能(Enable)
  - ▶ 部件:计数器、指令寄存器、指令解码器、状态寄存器、时钟发生器、微操作信号发生器

# CPU基本结构



# CPU性能指标

## ▶ 主频

- ▶ 主频也叫时钟频率, 用来表示CPU的运算、处理数据的速度.
- ▶ 主频和实际的运算速度存在一定的关系, 但并不是一个简单的线性关系.
- ▶ 主频 = 外频 × 倍频

## ▶ 前端总线(FSB)频率

- ▶ 前端总线(FSB)频率是直接影响CPU与内存直接数据交换速度.

# CPU性能指标

## ▶ CPU的位和字长

- ▶ 位：在数字电路和电脑技术中采用二进制,代码只有“0”和“1”,其中无论是“0”或是“1”在CPU中都是一“位”。
- ▶ 字长：CPU在单位时间内(同一时间)能一次处理的二进制数的位数叫字长。

# CPU性能指标

## ▶ 高速缓存

- ▶ CPU高速缓存(Cache)是用于减少处理器访问内存所需平均时间的部件。在金字塔式存储体系中它位于自顶向下的第二层,仅次于CPU寄存器。其容量远小于内存,但速度却可以接近处理器的频率。
- ▶ 缓存之所以有效,主要是因为程序运行时对内存的访问呈现局部性(Locality)特征。这种局部性既包括空间局部性(Spatial Locality),也包括时间局部性(Temporal Locality)。有效利用这种局部性,缓存可以达到极高的命中率。
- ▶ 高速缓存分为:一级缓存、二级缓存和三级缓存

# CPU指令集架构

- ▶ 从现阶段的主流体系结构讲, **指令集可分为复杂指令集(CISC)和精简指令集(RISC)两部分**(指令集共有四个种类)
- ▶ 从具体运用看, 如Intel的MMX(Multi Media Extended)、SSE(Streaming-Single instruction multiple data-Extensions)、SSE2、SSE3、SSE4系列和AMD的3DNow!等都是CPU的扩展指令集, 分别增强了CPU的多媒体、图形图象和Internet等的处理能力.

# CPU指令集架构

## ▶ 复杂指令集

- ▶ CISC = Complex Instruction Set Computer
- ▶ CISC的特点是指令数目多而复杂, 每条指令字长并不相等.

## ▶ 精简指令集

- ▶ RISC = Reduced Instruction Set Computer
- ▶ 对指令数目和寻址方式都做了精简, 使其实现更容易, 指令并行执进程度更好, 编译器的效率更高.

# CPU指令集架构

## ▶ IA-64

- ▶ IA-64(Intel Architecture-64 bit)架构是EPIC(Explicitly Parallel Instruction Computing)的64位架构. 通过将多条指令放入一个指令字, 有效的提高了CPU各个计算功能部件的利用效率, 提高了程序的性能.

## ▶ x86-64

- ▶ “x86-64”, 有时会简称为 “x64”, 是64位微处理器架构及其相应指令集的一种.

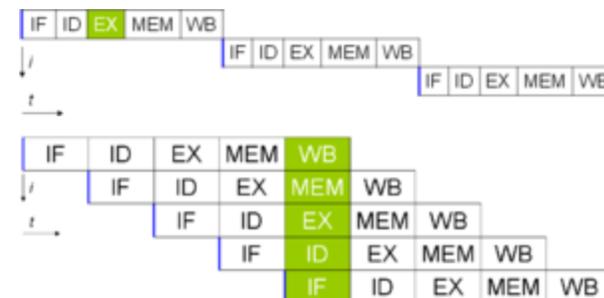
# CPU技术架构

## ▶ 流水线

- ▶ 流水线, 是现代计算机处理器中必不可少的部分, 是指将计算机指令处理过程拆分为多个步骤, 并通过多个硬件处理单元并行执行来加快指令执行速度. 其具体执行过程类似工厂中的流水线, 并因此得名.

## ▶ 并行计算

- ▶ 并行计算(Parallel computing)一般是指许多指令得以同时进行的计算模式.



# CPU技术架构

## ▶ 超线程(面向硬件)

- ▶ HT, Hyper-Threading

- ▶ 英特尔的HT技术是在CPU内部仅复制必要的资源、让CPU模拟成两个线程; 也就是一个实体核心, 两个逻辑线程, 在一单位时间内处理两个线程的工作, 模拟实体双核心、多线程运作.

## ▶ 多线程(面向软件)

- ▶ Multithreading

- ▶ 即便处理器只能运行一个线程, 操作系统也可以通过快速的在不同线程之间进行切换, 由于时间间隔很小, 来给用户造成一种多个线程同时运行的假象.

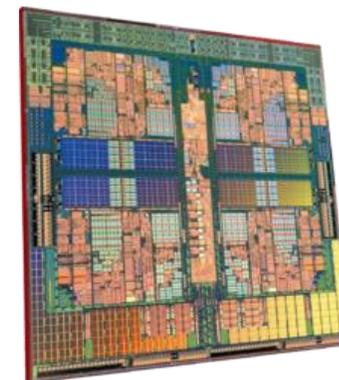
# CPU技术架构

## ▶ 多核心

- ▶ 双核中央处理器是在中央处理器芯片或封装中包含两个处理器核心,一般共用二级缓存. 现今使用双核心处理器的个人电脑已相当普遍. 另也有三核心、四核心处理器、六核心处理器等.

## ▶ 对称多处理结构

- ▶ SMP(Symmetric Multi-Processing)
- ▶ 指在一个计算机上汇集了一组处理器(多CPU), 各CPU之间共享内存子系统以及总线结构.



# 内存(主存储器)

- ▶ 内存(Memory)是一种利用半导体技术做成的电子设备,用来存储数据. 电子电路的数据是以二进制的方式存储,存储器的每一个存储单元称做记忆元.

# 内存的分类

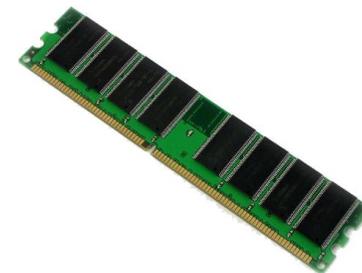
- ▶ 易失性存储器(Volatile memory)
  - ▶ 当电源供应中断后, 存储器所存储的数据便会消失的存储器
  - ▶ RAM(Random Access Memory, 随机访问存储器)
    - ▶ SRAM(Static random access memory, 静态随机访问存储器)
    - ▶ DRAM(Dynamic random access memory, 动态随机访问存储器)

# 内存的分类

- ▶ 非易失性存储器(Non-volatile memory)
  - ▶ 即使电源供应中断, 存储器所存储的数据并不会消失, 重新供电后, 就能够读取内存数据的存储器.
  - ▶ ROM(Read-only memory, 只读存储器)
  - ▶ PROM(Programmable read-only memory, 可编程只读存储器)
  - ▶ EPROM(Erasable programmable read only memory, 可擦可编程只读存储器)
  - ▶ EEPROM (Electrically erasable programmable read only memory, 可电擦可编程只读存储器)
  - ▶ Flash memory(闪存)

# 随机存储器(RAM)

- ▶ SRAM(静态随机存取存储器)
  - ▶ 以反相器交叉耦合构成的触发器作为存储单元. 不需要定时刷新, 读取速度快, 电路复杂度高, 成本较高. 常用于制造高速缓存(Cache).
- ▶ DRAM(动态随机存取存储器)
  - ▶ 主要的作用原理是利用电容内存储电荷的多少来代表一个二进制位(bit)是1还是0.
  - ▶ 由于在现实中电容会有漏电的现象, 导致电位差不足而使记忆消失, 因此除非电容经常周期性地充电, 否则无法确保记忆长存, 因此需要定时刷新



# 随机存储器(RAM)

## ▶ SRAM与DRAM的比较

- ▶ 与SRAM相比, DRAM的优势在于结构简单——每一个位的数据都只需一个电容跟一个晶体管来处理, 相比之下在SRAM上一个位通常需要六个晶体管. 正因这缘故, DRAM拥有非常高的密度, 单位体积的容量较高因此成本较低. 但相反的, DRAM也有访问速度较慢, 耗电量较大的缺点.

# 只读存储器(ROM)

- ▶ 只读存储器(Read-Only Memory)是一种半导体存储器, 其特性是一旦存储数据就无法再将之改变或删除, 且内容不会因为电源关闭而消失.
- ▶ ROM的种类
  - ▶ ROM(Read-Only Memory)
  - ▶ PROM(Programmable ROM)
  - ▶ EPROM(Erasable Programmable ROM)
  - ▶ OTPROM(One Time Programmable ROM)
  - ▶ EEPROM(Electrically Erasable Programmable ROM)
  - ▶ Flash Memory(闪存)

# BIOS和CMOS

## ▶ BIOS

- ▶ BIOS(Basic Input /Output System), 是加载在电脑硬件系统上的最基本的软件代码.
- ▶ 由于BIOS与硬件系统集成在一起, 所以有时候也被称为固件.

## ▶ CMOS

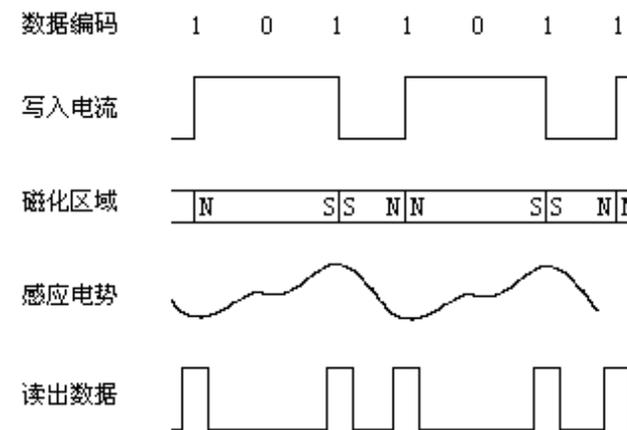
- ▶ CMOS(Complementary Metal-Oxide-Semiconductor)是一种集成电路制程.
- ▶ CMOS具有只有在晶体管需要切换启闭时才需耗能的优点, 因此非常省电且发热少, 系统的BIOS程序和参数信息都保存在CMOS中.

# EFI

- ▶ EFI(Unified Extensible Firmware Interface)
  - ▶ 可扩展固件接口(Unified Extensible Firmware Interface 或UEFI)是一种个人电脑系统规格,用来定义操作系统与系统固件之间的软件界面,为替代BIOS的升级方案.可扩展固件接口负责加电自检(POST)、联系操作系统以及提供连接操作系统与硬件的接口.

# 外存(辅存)——机械硬盘

- ▶ 硬盘(Hard Disk Drive, 简称HDD)是电脑上使用坚硬的旋转盘片为基础的非易失性(Non-Volatile)存储设备. 它在平整的磁性表面存储和检索数字数据. 信息通过离磁性表面很近的写头, 由电磁流来改变极性方式被电磁流写到磁盘上. 信息可以通过相反的方式回读.



# 机械硬盘参数

## ▶ 容量

- ▶ 目前硬盘的容量有250GB、320GB、500GB、640GB、750GB、1TB、1.5TB、2TB、2.5TB、3TB等多种规格

## ▶ 转速

- ▶ 硬盘每分钟旋转的圈数,单位是rpm(每分钟的转动数),有5400rpm、7200rpm、10000rpm、15000rpm等几种规格

## ▶ 缓存

- ▶ 主要有2MB、8MB、16MB、32MB、64MB等规格

## ▶ 平均寻道时间

- ▶ 单位是ms(毫秒),有5.2ms、8.5ms、8.9ms、12ms等规格

# 机械硬盘结构

## ▶ 磁道

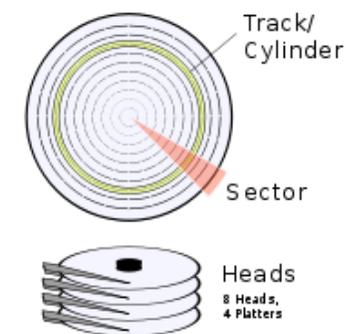
- ▶ 当磁盘旋转时, 磁头若保持在一个位置上, 则每个磁头都会在磁盘表面划出一个圆形轨迹, 这些圆形轨迹就叫做磁道(Track).

## ▶ 柱面

- ▶ 在有多个盘片构成的盘组中, 由处于同一半径的磁道组成的一个圆柱面(Cylinder).

## ▶ 扇区

- ▶ 磁盘上的每个磁道被等分为若干个弧段, 这些弧段便是硬盘的扇区(Sector). 硬盘的第一个扇区, 叫做引导扇区.



# 固态硬盘

- ▶ 固态硬盘(Solid State Drive, 简称SSD, 固态驱动器)是一种基于永久性存储器, 如闪存, 或非永久性存储器, 同步动态随机存取存储器(SDRAM)的计算机外部存储设备.
- ▶ 虽然在固态硬盘中已经没有可以旋转的盘状结构, 但是依照人们的命名习惯, 这类存储器仍然被称为“硬盘”.
- ▶ 新一代的固态硬盘普遍采用SATA-3接口.



# 固态硬盘

## ▶ 固态硬盘的优点

- ▶ 和传统硬盘相比,固态硬盘具有高速、低功耗、无噪音、抗震动、低热量的特点. 这些特点不仅使得数据能更加安全地得到保存,而且也延长了靠电池供电的设备的连续运转时间.

## ▶ 固态硬盘的缺点

- ▶ 目前固态硬盘普及的最大问题仍然是成本和写入次数,因此只有小容量的固态硬盘的价格能够被大多数人所承受.
- ▶ 由于闪存都有一定的写入寿命、而且寿命届满后数据会读不出来,因此成为大众接受固态硬盘的另一个障碍.

# 硬盘接口

- ▶ PATA(Parallel Advanced Technology Attachment)
  - ▶ PATA是用传统的40-pin并行数据线连接主板与硬盘的,外部接口速度最大为133MB/s.
- ▶ SATA(Serial ATA)
  - ▶ 在数据传输上这一方面, SATA的速度比以往更加快捷, 并支持热插拔, 使电脑运作时可以插上或拔除硬件. 另一方面, SATA总线具备了比以往更强的纠错能力, 能对传输指令进行检查, 如果发现错误会自动矫正, 提高了数据传输的可靠性.
  - ▶ SATA分别有SATA 1.5Gbit/s、SATA 3Gbit/s和SATA 6Gbit/s三种规格.
- ▶ eSATA(External Serial ATA)
  - ▶ 为面向外接驱动器而制定的Serial ATA 1.0a的扩展规格, 传输速度可以达到现在主流的USB 2.0的传输速度的2倍以上.

# 硬盘接口

- ▶ SCSI(Small Computer System Interface)
  - ▶ 小型计算机系统接口是一种用于计算机及其周边设备之间(硬盘、软驱、光驱、打印机、扫描仪等)系统级接口的独立处理器标准. 最大部分的应用是在存储设备上(例如硬盘、磁带机).
- ▶ SAS(Serial Attached SCSI)
  - ▶ 串行SCSI是由并行SCSI物理存储接口演化而来, 与并行方式相比, 串行方式提供更快速的通信传输速度以及更简易的配置.

# RAID

- ▶ 独立磁盘冗余数组(RAID, Redundant Array of Independent Disks)简称硬盘阵列, 其基本思想就是把多个相对便宜的硬盘组合起来, 成为一个硬盘阵列组, 使性能达到甚至超过一个价格昂贵、容量巨大的硬盘.
- ▶ 根据选择的版本不同, RAID比单颗硬盘有以下几个或多个方面的好处:增强数据集成度, 增强容错功能, 增加处理量或容量, 看起来就像一个单独的硬盘或逻辑存储单元. 分为RAID-0, RAID-1, RAID-1E, RAID-5, RAID-6, RAID-7, RAID-10, RAID-50.



# 外存(辅存)——光盘

- ▶ 光盘于1965年由美国发明,它是用激光扫描的记录和读出方式保存信息的一种介质. 大约在1990年代左右时开始普及,具有存放大量数据的特性.
- ▶ 光盘类型:
  - ▶ 红外激光
    - ▶ CD(700MB)
  - ▶ 红色激光
    - ▶ DVD(单面单层4.7GB, 单面双层:9.4GB; 双面单层:8.5GB, 双面双层:17GB)
  - ▶ 蓝色激光
    - ▶ HD DVD(High Definition DVD)(单面单层:15GB...)
    - ▶ BD(Blue-ray Disk)(单层:25 GB; 双层:50 GB...)



# 显卡(Graphics Card)

- ▶ 显卡是个人电脑最基本组成部分之一. 显卡的用途是将计算机系统所需要的显示信息进行转换驱动显示器, 并向显示器提供行扫描信号, 控制显示器的正确显示.
- ▶ 总线接口类型
  - ▶ ISA显卡
  - ▶ PCI显卡
  - ▶ AGP显卡
  - ▶ PCI Express显卡



# 图形处理器(GPU)

- ▶ GPU(Graphics Processing Unit)是一种专门用来处理在个人电脑、工作站或游戏机上那些图像运算工作的微处理器。GPU可以集成在显卡上(独立显卡),或者直接集成到CPU上(核心显卡).
- ▶ GPU是显示卡的“心脏”,也就相当于CPU在电脑中的作用,它决定了该显卡的档次和大部分性能.
- ▶ GPU生产厂商
  - ▶ Intel HD Graphics
  - ▶ Nvidia
  - ▶ ATI(被AMD收购)

# 总线(Bus)

- ▶ 总线(Bus)是指计算机组件间规范化的交换数据(Data)的方式, 即以一种通用的方式为各组件提供数据传送和控制逻辑.
- ▶ 这些线路在同一时间内都仅能负责传输一个比特. 因此, 必须同时采用多条线路才能传送更多数据, 而总线可同时传输的数据数就称为宽度(Width), 以比特为单位, 总线宽度愈大, 传输性能就愈佳.
- ▶ 总线的带宽(即单位时间内可以传输的总数据数)为: 总线带宽 = 频率 × 宽度(Bytes/sec)

# 总线(Bus)

## ▶ 总线的分类:

### ▶ 片内总线(Chip Bus)

▶ 又称前端总线, 是把各种不同的芯片连接在一起构成特定功能模块的信息传输通路

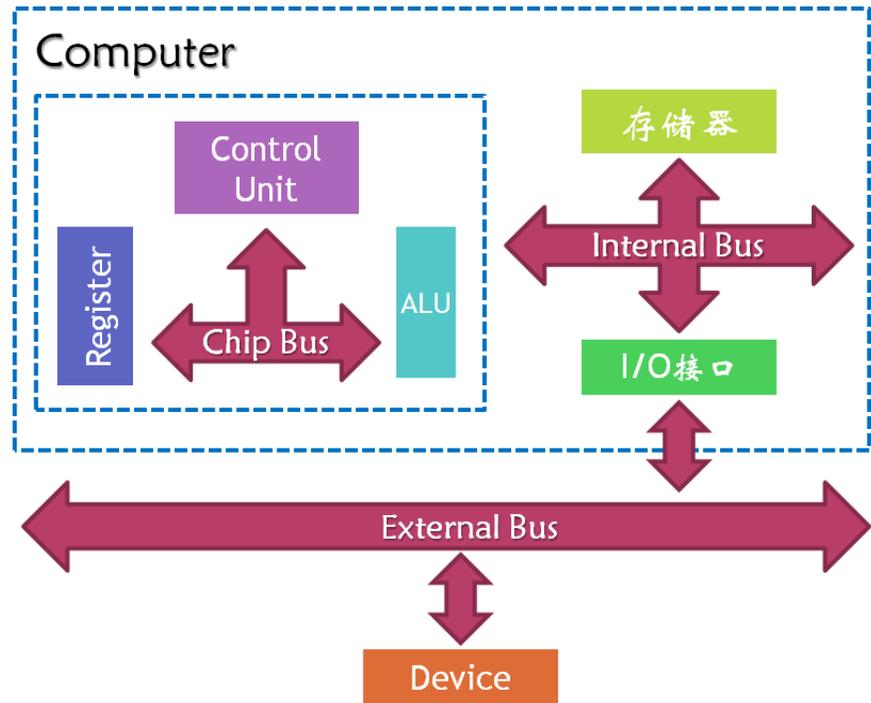
### ▶ 内部总线(Internal Bus)

▶ 又称系统总线, 是微机系统中各插件(模块)之间的信息传输通路

### ▶ 外部总线(External Bus)

▶ 又称通信总线, 是微机系统之间或微机系统与其他系统之间信息传输的通路

# 总线(Bus)



# 总线(Bus)

- ▶ 内部总线的种类:

- ▶ 数据总线(Data Bus)
- ▶ 地址总线(Address Bus)
- ▶ 控制总线(Control Bus)

- ▶ 常见的总线类型

- ▶ PCI, PCI Express, USB, IEEE 1394, PS/2, SATA, etc.

# 计算机网络

# 计算机网络概述

- ▶ 计算机网络是利用通信设备和线路将地理位置不同的、功能独立的多个计算机系统连接起来,以功能完善的网络软件实现网络的硬件、软件及资源共享和信息传递的系统. 简单的说即连接两台或多台计算机进行通信的系统.

# 计算机网络分类

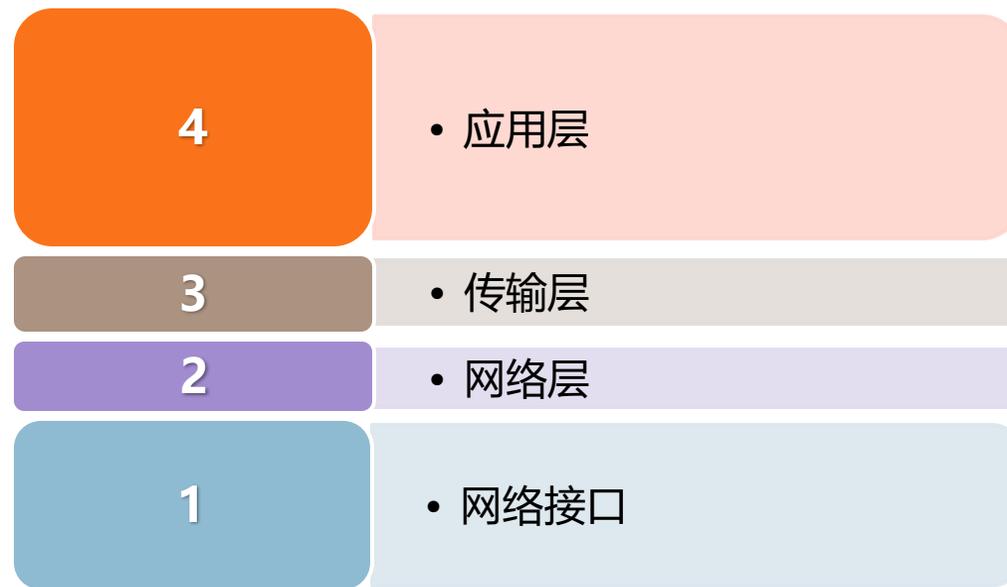
- ▶ 按覆盖范围分类:
  - ▶ 局域网(Local Area Network, LAN)
  - ▶ 城域网 (Metropolitan Area Network, MAN)
  - ▶ 广域网(Wide Area Network, WAN)
- ▶ 按不同节点之间的功能关系分类:
  - ▶ 客户/服务器(C/S或B/S)
  - ▶ 对等结构(P2P)

# 网络协议分层

## OSI参考模型



## TCP/IP 模型



# 网络协议

## 应用层

- DHCP、DNS、FTP、HTTP、IMAP4、POP3、SMTP、SNMP、SSH、TELNET、RPC.....

## 传输层

- TCP、UDP、TLS、DCCP、SCTP、RSVP、PPTP、OSPF.....

## 网络层

- IP、ARP、RARP、ICMP、ICMPv6、IGMP、IS-IS.....

## 物理层

- Wi-Fi(IEEE 802. 11)、WiMax(IEEE 802. 16) 、Ethernet.....
- 调制解调器、光导纤维、同轴电缆、双绞线.....



# 计算机网络设备

- ▶ 计算机网络组成硬件
  - ▶ 网络适配器 <Network Interface Card>
  - ▶ 中继器 <Repeater>
  - ▶ 集线器 <Hub>
  - ▶ 桥接器 <Bridge>
  - ▶ 交换机 <Switch>
  - ▶ 路由器 <Router>
- ▶ (\*) Difference between Switch and Router.

# MAC地址

- ▶ MAC地址(Media Access Control Address), 媒体访问控制地址, 或称为硬件地址, 是用来定义网络设备的位置的. 在OSI模型中, 第三层网络层负责IP地址, 第二层数据链路层则负责MAC地址. 一个主机会有一个IP地址, 而每个网络位置会有一个专属于它的MAC地址.
- ▶ MAC地址共48位(6个字节), 以十六进制表示. 前24位由IEEE等组织决定如何分配, 后24位由实际生产该网络设备的厂商自行指定.
- ▶ Example: 60-67-20-02-6E-C6

# IP地址

- ▶ IP地址 ( Internet Protocol Address ) 是一种在Internet上的给主机编址的方式, 也称为网际协议地址.
- ▶ 常见的IP地址, 分为IPv4与IPv6两大类.
- ▶ IP是一种无连接的协议, 操作在使用分组交换的链路层(如以太网)上. 此协议会尽最大努力交付分组, 意即它不保证任何分组均能送达目的地, 也不保证所有分组均按照正确的顺序无重复地到达. 这些方面是由上层的传输协议(如传输控制协议)处理的.

# IPv4地址

- ▶ 由32位二进制数组成, 为便于使用, 常以XXX. XXX. XXX. XXX形式表现, 每组XXX代表[0, 255]的10进制数.
- ▶ 例如218. 108. 45. 154. 地址可分为A, B, C, D, E五大类, 其中E类属于特殊保留地址.
- ▶ 在互联网中, IP地址是唯一的. IP v4技术可能使用的IP地址最多可有约42亿个.乍看可能觉得很难会用尽, 但由于早期编码上的问题, 使很多编码实际上被丢空或不能使用.
- ▶ 随着互联网的快速成长, IPv4的42亿个地址最终于2011年2月3日用尽.

# A类IP地址

## ▶ A类IP地址

- ▶ A类IP地址就由1字节的网络地址和3字节主机地址组成, 网络地址的最高位必须是“0”.
- ▶ A类IP地址中网络的标识长度为7位, 主机标识的长度为24位, A类网络地址数量较少, 可以用于主机数达1600多万台的大型网络.
- ▶ A类IP地址范围1.0.0.1-126.255.255.254
- ▶ A类IP地址的子网掩码为255.0.0.0, 每个网络支持的最大主机数为 $2^{24}-2=16777216$ 台

# B类IP地址

## ▶ B类IP地址

- ▶ B类IP地址就由2字节的网络地址和2字节主机地址组成, 网络地址的最高位必须是“10”.
- ▶ B类IP地址中网络的标识长度为14位, 主机标识的长度为16位, B类网络地址适用于中等规模的网络, 每个网络所能容纳的计算机数为65534台.
- ▶ B类IP地址的地址范围128.0.0.1-191.255.255.254
- ▶ B类IP地址的子网掩码为255.255.0.0, 每个网络支持的最大主机数为 $256^2-2=65534$ 台

# C类IP地址

## ▶ C类IP地址

- ▶ C类IP地址就由3字节的网络地址和1字节主机地址组成, 网络地址的最高位必须是“110”.
- ▶ C类IP地址中网络的标识长度为21位, 主机标识的长度为8位, C类网络地址数量较多, 适用于小规模的网络, 每个网络最多只能包含254台计算机.
- ▶ C类IP地址范围192.0.0.1-223.255.254.254
- ▶ C类IP地址的子网掩码为255.255.255.0, 每个网络支持的最大主机数为 $256-2=254$ 台

# D类和E类IP地址

## ▶ D类IP地址

- ▶ D类地址用于多点播送, 第一个字节以“1110”开始, 第一个字节的数字范围为224~239, 是多点播送地址, 用于多目的地信息的传输, 和作为备用.  
D类地址: 属多点播送地址 224. 0. 0. 0 ~ 239. 255. 255. 255

## ▶ E类IP地址

- ▶ 以“1111”开始, 为将来使用保留. 它的第一字节的范围是240~255, 主要用于Internet试验和开发.

# IPv4专用网络

- ▶ A类保留地址
  - ▶ 地址范围:10. 0. 0. 0–10. 255. 255. 255
  - ▶ 地址数量:16, 777, 216
- ▶ B类保留地址
  - ▶ 地址范围:172. 16. 0. 0–172. 31. 255. 255
  - ▶ 地址数量:1, 048, 576
- ▶ C类保留地址
  - ▶ 地址范围:192. 168. 0. 0–192. 168. 255. 255
  - ▶ 地址数量:65, 536

# IPv6地址

- ▶ IPv6(Internet Protocol version 6)是被指定为IPv4继任者的下一代互联网协议版本. IPv6具有比IPv4大得多的地址空间, 这是因为IPv6使用了128位的地址, 而IPv4只用32位. 因此新增的地址空间支持 $2^{128}$ 个地址.
- ▶ 在很多场合, IPv6地址由两个逻辑部分组成: 一个64位的网络前缀和一个64位的主机地址, 主机地址通常根据物理地址自动生成, 叫做EUI-64(或者64-位扩展唯一标识).
- ▶ Example: 2001:da8:d805:a203:956f:c651:e62:6a0

# IPv6地址格式

- ▶ IPv6在某些条件下可以省略:
  - ▶ 规则1：每项数字前导的0可以省略, 省略后前导数字仍是0则继续
  - ▶ 例如 以下IPv6地址是等价的
    - ▶ 2001:0DB8:02de:0000:0000:0000:0000:0e13
    - ▶ 2001:DB8:2de:0000:0000:0000:0000:e13
    - ▶ 2001:DB8:2de:000:000:000:000:e13
    - ▶ 2001:DB8:2de:00:00:00:00:e13 2001:DB8:2de:0:0:0:0:e13

# IPv6地址格式

- ▶ IPv6在某些条件下可以省略:
  - ▶ 规则2：若有连贯的0000的情形出现, 可以用双冒号 "::" 代替.
  - ▶ 例如 以下IPv6地址是等价的
    - ▶ 2001:0DB8:0000:0000:0000:0000:1428:57ab
    - ▶ 2001:0DB8:0000:0000:0000::1428:57ab
    - ▶ 2001:0DB8:0:0:0:0:1428:57ab
    - ▶ 2001:0DB8:0::0:1428:57ab
    - ▶ 2001:0DB8::1428:57ab

# IPv6地址格式

- ▶ 不过请注意有的情形下省略是非法的, 例如这个IPv6是非法的:
  - ▶ 2001::25de::cade
  - ▶ 因为它有可能是下种情形之一, 造成无法推断:
    - ▶ 2001:0000:0000:0000:0000:25de:0000:cade
    - ▶ 2001:0000:0000:0000:25de:0000:0000:cade
    - ▶ 2001:0000:0000:25de:0000:0000:0000:cade
    - ▶ 2001:0000:25de:0000:0000:0000:0000:cade

# IPv4映射地址

- ▶ 如果一个地址实际上是IPv4的地址, 后32位可以用10进制数表示; 因此:
  - ▶ ffff:218. 108. 45. 154等价于::ffff:da6c:2d9a
- ▶ IPv4位址可以很容易的转化为IPv6格式.
  - ▶ IPv4的一个地址为218.108. 45. 154
  - ▶ 它可以被转化为0000:0000:0000:0000:0000:da6c:2d9a或者::da6c:2d9a.
  - ▶ 同时, 还可以使用混合符号(IPv4-compatible address), 则地址可以为::218.108. 45. 154

# TCP和UDP协议

## ▶ TCP协议

- ▶ 传输控制协议(Transmission Control Protocol, TCP)是一种面向连接的、可靠的、基于字节流的传输层(Transport layer)通信协议.
- ▶ TCP使用序号, 对收到的TCP报文段进行排序以及检测重复的数据; 使用确认和计时器来检测和纠正丢包或延时.

## ▶ UDP协议

- ▶ 用户数据报协议(User Datagram Protocol, UDP)是一个简单的面向数据报的传输层协议.
- ▶ 由于缺乏可靠性且属于非连接导向协定, UDP应用一般必须允许一定量的丢包、出错和复制.

# 计算机网络拓扑结构

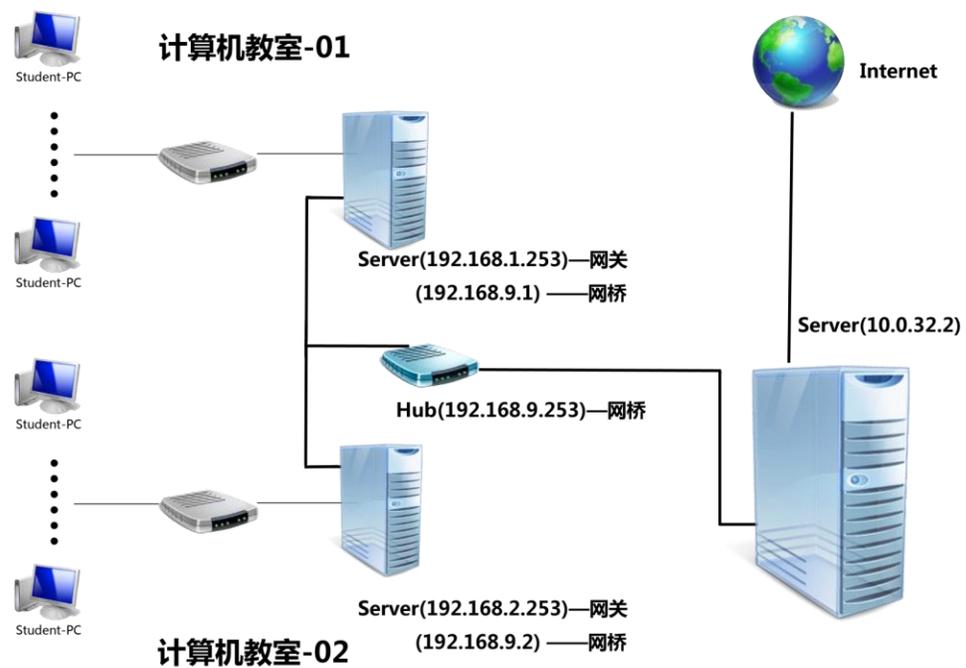
## ▶ 拓扑结构概述:

- ▶ 计算机网络拓扑结构是指网络中各个站点相互连接的形式, 在局域网中明确一点讲就是文件服务器、工作站和电缆等的连接形式.

## ▶ 计算机网络的拓扑(Topologies)结构可分为

- ▶ 总线拓扑(Bus Network)
- ▶ 星型拓扑(Star Network)
- ▶ 环型拓扑(Ring Network)
- ▶ 全互联拓扑(Mesh Network)
- ▶ 树状拓扑(Tree Network)

# 校园网络拓扑结构



# 子网掩码

- ▶ 子网掩码(Subnet Mask)是一种用来指明一个IP地址的哪些位标识的是主机所在的子网以及哪些位标识的是主机的位掩码.
- ▶ 子网掩码只有一个作用,就是将某个IP地址划分成网络地址和主机地址两部分.
- ▶ 在IPv4中,就是点分四组表示法(四个取值从0到255的数字由点隔开,比如255. 128. 0. 0)

# 子网掩码的计算

- ▶ 已知子网掩码确定主机数目
  - ▶ 计算公式是 $2^m$ , 其中, 我们可以把 $m$ 看作是后面的多少个0.
- ▶ Question:
  - ▶ 计算子网掩码为255.255.255.0子网的主机数目.
- ▶ Answer:
  - ▶ 将255.255.255.0转成二进制得: 11111111.11111111.11111111.00000000
  - ▶ 末尾有8个零,  $m=8$
  - ▶ 该子网理论上可容纳 $2^8$ 台计算机, 实际上是254台

# 子网掩码的计算

▶ Question:

- ▶ 计算子网掩码为255.255.248.0子网的主机数目.

▶ Answer:

- ▶ 将255.255.255.0转成二进制得: 11111111.11111111.11111000.00000000
- ▶ 末尾有11个零,  $m=11$
- ▶ 该子网理论上可容纳 $2^{11}$ 台计算机, 实际上是2046台

# 子网掩码的计算

- ▶ 已知主机数目确定子网掩码
  - ▶ 计算公式是 $m = \log_2 N$ , 其中, 我们可以把 $m$ 看作是后面的多少个0.
- ▶ Question:
  - ▶ 已知子网中中共有254台计算机, 则子网掩码至少为多少?
- ▶ Answer:
  - ▶  $\because 256 = 2^8$ , 即 $m = 8$
  - ▶ 即子网掩码最后有8个零, 即 11111111.11111111.11111111.00000000
  - ▶ 转换成十进制得255. 255. 255. 0

# 子网掩码的计算

▶ Question:

- ▶ 已知子网中共有336台计算机, 则子网掩码至少为多少?

▶ Answer:

- ▶  $\because 2^8 = 256 < 336, 2^9 = 512 > 336, \therefore m = 9$
- ▶ 即子网掩码最后有9个零, 即 11111111.11111111.11111110.00000000
- ▶ 转换成十进制得255.255.254.0

# 子网掩码的计算

- ▶ 子网掩码是用来判断任意两台计算机的IP地址是否属于同一子网络的根据.
  - ▶ 最为简单的理解就是两台计算机各自的IP地址与子网掩码进行And运算后, 如果得出的结果是相同的, 则说明这两台计算机是处于同一个子网络上的, 可以进行直接的通讯.

# 子网掩码的计算

- ▶ IP地址A:192. 168. 0. 1; 子网掩码:255. 255. 255. 0
- ▶ 转换成二进制:
  - ▶ IP地址:11000000.10101000.00000000.00000001
  - ▶ 子网掩码:11111111.11111111.11111111.00000000
  - ▶ And运算结果11000000.10101000.00000000.00000000
- ▶ 转换成十进制:
  - ▶ 192. 168. 0. 0

# 子网掩码的计算

- ▶ IP地址B:192. 168. 0. 254; 子网掩码:255. 255. 255. 0
- ▶ 转换成二进制:
  - ▶ IP地址:11000000. 10101000. 00000000. 11111110
  - ▶ 子网掩码:11111111. 11111111. 11111111. 00000000
  - ▶ And运算结果11000000. 10101000. 00000000. 00000000
- ▶ 转换成十进制:
  - ▶ 192. 168. 0. 0

# 域名(Domain Name)

## ▶ 域名的起源

- ▶ IP地址是Internet主机的作为路由寻址用的数字型标识, 人不容易记忆. 因而产生了域名(Domain Name)这一种字符型标识.

## ▶ 域名的概述

- ▶ 域名(Domain Name), 是由一串用点分隔的名字组成的Internet上某一台计算机或计算机组的名称, 用于在数据传输时标识计算机的电子方位(有时也指地理位置).
- ▶ 例如:zjhzxhz.com作为一个域名, 与IP地址96.44.131.251相对应.

# DNS(Domain Name System)

- ▶ DNS(Domain Name System)
  - ▶ DNS是因特网的一项核心服务,它作为可以将域名和IP地址相互映射的一个分布式数据库,能够使人更方便的访问互联网,而不用去记住能够被机器直接读取的IP地址数串.
- ▶ 域名服务器(Domain Name Server)
  - ▶ 域名服务器(Domain Name Server, 简称Name Server)实际上就是装有域名系统的主机.它是一种能够实现域名解析的分层结构数据库.

# 网络攻击

## ▶ 网络攻击的分类

- ▶ 主动攻击: 包含攻击者访问所需要信息的故意行为

- ▶ 被动攻击:

- ▶ 窃听: 包括键击记录、网络监听、非法访问数据、获取密码文件

- ▶ 欺骗: 包括获取口令、恶意代码、网络欺骗

- ▶ 拒绝服务(DOS): 包括导致异常型、资源耗尽型、欺骗型

- ▶ 数据驱动攻击: 包括缓冲区溢出、格式化字符串攻击、输入验证攻击、同步漏洞攻击、信任漏洞攻击

# 计算机病毒

## ▶ 计算机病毒概述

- ▶ 计算机病毒(Computer Virus)是一种在用户不知情或批准下,能自我复制及运行的电脑程序,电脑病毒往往会影响受感染电脑的正常运作.
- ▶ 在计算机科学里,电脑病毒是类似生物病毒一样的程序,它会复制自己并传播到其他宿主,并对宿主造成损害.宿主也是程序,通常是操作系统,从而进一步传染到其他程序、其他的电脑.电脑病毒在传播期间一般会隐蔽自己,由特定的条件触发,并开始产生破坏.

# 网络应用程序

## ▶ 防病毒软件

- ▶ Kaspersky、McAfee、NOD32、Norton、360、瑞星杀毒软件、金山毒霸.....

## ▶ 即时通信软件

- ▶ ICQ、Skype、QQ、Google Talk、UC.....

## ▶ 网页浏览器

- ▶ Internet Explorer、Mozilla Firefox、Google Chrome、Opera、Safari.....

## ▶ 其他应用软件

- ▶ FileZilla, FlashFXP、迅雷、Foxmail.....

# 网络新概念

## ▶ HTTP 2.0

- ▶ HTTP/2 (originally named HTTP/2.0) is the second major version of the HTTP network protocol used by the World Wide Web. It is based on SPDY.
- ▶ HTTP/2 was developed by the Hypertext Transfer Protocol working group (httpbis, where bis means "repeat" or "twice") of the Internet Engineering Task Force.
- ▶ HTTP/2 is the first new version of HTTP since HTTP 1.1, which was standardized in RFC 2068 in 1997.

# 网络新概念

## ▶ CDN

- ▶ A content delivery network (CDN) is a system of distributed servers (network) that deliver webpages and other Web content to a user based on the geographic locations of the user, the origin of the webpage and a content delivery server.

## ▶ Docker

- ▶ Docker is an open-source project that automates the deployment of applications inside software containers, by providing an additional layer of abstraction and automation of operating-system-level virtualization on Linux, Mac OS and Windows.

# 网络新概念

## ▶ HTML 5

- ▶ HTML5 is a core technology markup language of the Internet used for structuring and presenting content for the World Wide Web. As of October 2014 this is the final and complete fifth revision of the HTML standard of the World Wide Web Consortium (W3C). The previous version, HTML 4, was standardized in 1997.

## ▶ 数据挖掘(Data Mining)

- ▶ Data mining is the analysis of historical business activities, stored as static data in data warehouse databases, to reveal hidden patterns and trends.

# 网络新概念

## ▶ 云计算(Cloud Computing)

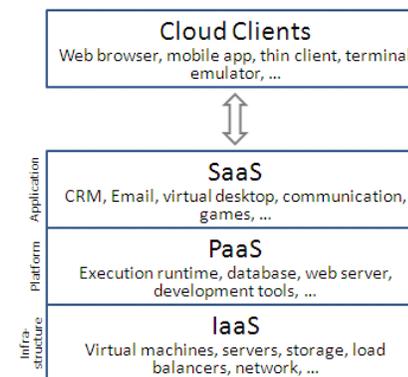
▶ Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network (typically the Internet).

▶ Service Models:

▶ IaaS, PaaS, SaaS

▶ Examples:

▶ Amazon AWS, Microsoft Azure, 阿里云, etc.



# 计算机信息数字化

# 数据与信息

## ▶ 数据

- ▶ 数据的概念: 数字、文字、图画、声音和活动图象
- ▶ 计算机系统中的数据一般以二进制编码形式出现的
- ▶ 数据区分为数值型数据和非数值型数据

## ▶ 信息

- ▶ 信息的概念: 信息是对人有用的数据

## ▶ 数据和信息的区别

- ▶ 数据包含信息
- ▶ 数据处理之后产生结果为信息
- ▶ 信息具有相对性, 时效性

# 媒体

## ▶ 媒体的概念

- ▶ 媒体又称媒介、媒质,是承载信息的载体.

## ▶ 媒体的分类

- ▶ 感觉媒体:视觉、听觉、味觉、嗅觉、触觉
- ▶ 表示媒体:文字、声音、图形、图像、视频
- ▶ 存储媒体:磁盘、光盘、光盘、内存
- ▶ 显示媒体:显示器、打印机、投影仪、音响
- ▶ 传输媒体:双绞线、同轴线缆、光纤

# 多媒体

## ▶ 声音

### ▶ 常见的音频格式

▶ wma, wav, mp3, aac, rm...

### ▶ 数据量计算

▶ 数据量 = 采样频率 × 量化位数 × 声道数 / 8 (Byte)

# 多媒体

## ▶ 图像

### ▶ 常见的图形格式

▶ jpg, bmp, png, gif, tif, tiff..

### ▶ 数据量计算

▶ RGB图像:数据量 = 图像长度 × 图像宽度 × 3 (Byte)

▶ 黑白图像:数据量 = 图像长度 × 图像宽度 / 8 (Byte)

# 多媒体

- ▶ 视频

- ▶ 常见的音频格式

- ▶ wmv, asf, mp4, avi, mpg, mov, rmvb, flv...

# 数制及其转换

## ▶ 数制的概念

▶ 即进位计数制, 是指计数的方法, 即采用一组计数符号(称为数符或数码)的组合来表示任意一个数的方法. 在进位计数法中, 数码序列中相同的一个数码所表示的数值大小与它在该数码序列中的位置有关. 无论使用何种进制, 他们都包含两个要素:

### ▶ 基数:

▶ 数制所使用的数码个数称为“基数”。

▶ 如十进制的基数是10。

### ▶ 位权

▶ 某数制每一位所具有的值称为“位权”。

▶ 例如  $101 = 1 \times 10^2 + 0 \times 10^1 + 2 \times 10^0$

# 计算机的数制

- ▶ 十进制(Decimal Notation)
  - ▶ 逢十进一
  - ▶ 0~9共十个数字来表示(基数为10)
- ▶ 二进制(Binary Notation)
  - ▶ 逢二进一
  - ▶ 使用两种不同的数字符号,即0或1(基数为2)
- ▶ 八进制(Octal Notation)
  - ▶ 逢八进一
  - ▶ 每位可取0~7中的任意一个数字(基数为8)
- ▶ 十六进制(Hexadecimal Notation)
  - ▶ 逢十六进一
  - ▶ 0~9这十个数由数字0~9来表示,10~15这六个数由英文字母A~F表示(基数为16)

# 计算机的数制

十进制	二进制	八进制	十六进制
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5
6	0110	6	6
7	0111	7	7

十进制	二进制	八进制	十六进制
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	10

# 计算机运算基础

## ▶ 计算机使用二进制的原因

### ▶ 容易表示

- ▶ 二进制数只有“0”和“1”两个基本符号,易于用两种对立的物理状态表示.
- ▶ 可用“1”表示电灯开关的“闭合”状态,用“0”表示“闭合”状态;晶体管的导通表示“1”,截止表示“0”;电容器的充电和放电.

### ▶ 运算简单

- ▶ 二进制数的算术运算特别简单,加法和乘法仅各有3条运算规则
- ▶ 运算时不易出错
- ▶ 二进制数的“1”和“0”正好可与逻辑值“真”和“假”相对应,这样就为计算机进行逻辑运算提供了方便.

# 逻辑代数基础

## ▶ 逻辑代数概述

- ▶ 逻辑代数起源于英国数学家乔治·布尔(George Boole)于1849年创立的布尔代数.
- ▶ 逻辑代数是一个由逻辑变量集、常量0和1及 “与” 、 “或” 、 “非” 三种运算所构成的代数系统.

# 二进制的逻辑运算

## ▶ 逻辑与(And)

### ▶ 定义:

▶ 逻辑与是两个逻辑变量的一种运算, 经常是两个命题的运算. 它满足: 当且仅当其两个变量的真值都为真时, 其结果为真.

### ▶ 位运算:

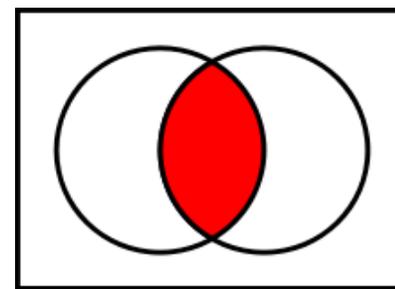
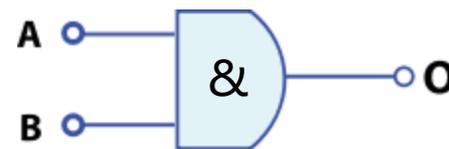
▶  $0 \wedge 0 = 0$

▶  $0 \wedge 1 = 0$

▶  $1 \wedge 0 = 0$

▶  $1 \wedge 1 = 1$

▶  $1100 \wedge 1010 = 1000$



# 二进制的逻辑运算

## ▶ 逻辑或(Or)

### ▶ 定义:

▶ 逻辑或是两个逻辑变量的一种运算, 经常是两个命题的运算. 它满足: 当且仅当其两个变量的真值都为假时, 其结果为假.

### ▶ 位运算:

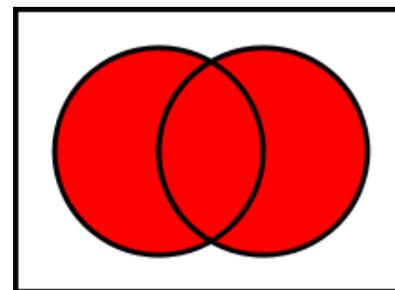
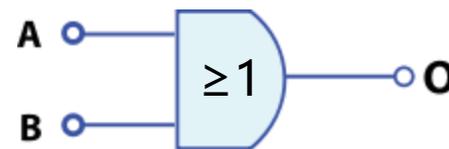
▶  $0 \vee 0 = 0$

▶  $0 \vee 1 = 1$

▶  $1 \vee 0 = 1$

▶  $1 \vee 1 = 1$

▶  $1100 \vee 1010 = 1110$



# 二进制的逻辑运算

## ▶ 逻辑非(Not)

### ▶ 定义:

▶ 逻辑非是布尔代数中一种一元运算. 它的运算结果是将运算元的真值取反.

### ▶ 命题 A 的非可以有几种写法

▶ Not A

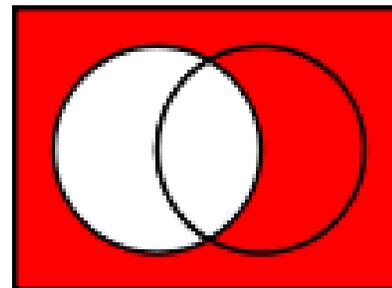
▶  $\sim A$  /  $\neg A$

▶ !A

### ▶ 位运算:

▶  $\neg 0 = 1$

▶  $\neg 1 = 0$



# 二进制的逻辑运算

## ▶ 逻辑异或(Xor)

### ▶ 定义:

▶ 逻辑异或是两个逻辑变量的一种运算, 经常是两个命题的运算. 它满足: 当且仅当其两个变量的真值都为假时, 其结果为假.

### ▶ 位运算:

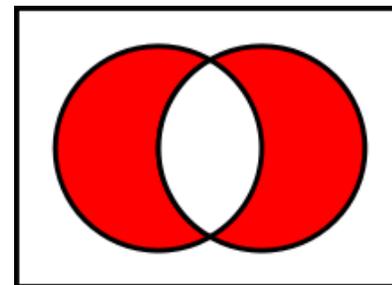
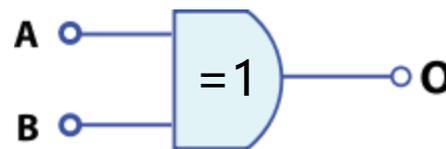
▶  $0 \vee 0 = 0$

▶  $0 \vee 1 = 1$

▶  $1 \vee 0 = 1$

▶  $1 \vee 1 = 0$

▶  $1100 \vee 1010 = 0110$



# 二进制的算术运算

## ▶ 四则运算法则

### ▶ 加法

▶  $00 + 00 = 00, 00 + 01 = 01, 01 + 00 = 01, 01 + 01 = 10$

### ▶ 减法

▶  $0 - 0 = 0, 1 - 0 = 1, 1 - 1 = 0, 10 - 1 = 1$

### ▶ 乘法

▶  $0 \times 0 = 0, 0 \times 1 = 0, 1 \times 0 = 0, 1 \times 1 = 1$

### ▶ 除法

▶  $0 \div 1 = 0, 1 \div 1 = 1$

# 进制间的转换

## ▶ 二、八、十六进制转换成十进制

### ▶ 二进制转换为十进制

$$\begin{aligned} \text{▶ } (1010101)_2 &= 1 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 64 + 16 + 4 + 1 = (85)_{10} \end{aligned}$$

### ▶ 八进制转换为十进制

$$\text{▶ } (34.6)_8 = (3 \times 8^1 + 4 \times 8^0 + 6 \times 8^{-1})_{10} = (24 + 4 + 0.75)_{10}$$

### ▶ 十六进制转换为十进制

$$\begin{aligned} \text{▶ } (16A.B)_{16} &= (1 \times 16^2 + 6 \times 16^1 + 10 \times 16^0 + 11 \times 16^{-1})_{10} \\ &= (256 + 96 + 10 + 0.69)_{10} = (362.69)_{10} \end{aligned}$$

# 进制间的转换

- ▶ 十进制数转换成二进制数
  - ▶ 纯整数部分的转换: 除2取余
  - ▶ 将 $(19)_{10}$ 转换为二进制数
  - ▶  $\therefore (19)_{10} = (10011)_2$



# 进制间的转换

- ▶ 十进制数转换成二进制数
  - ▶ 纯小数部分的转换:乘2取整
  - ▶ 将 $(0.125)_{10}$ 转换为二进制数
  - ▶  $\therefore (0.125)_{10} = (0.001)_2$

$$\begin{array}{r} 0.125 \quad (\text{高位}) \\ \times \quad 2 \\ \hline 0.250 \\ \times \quad 2 \\ \hline 0.500 \\ \times \quad 2 \\ \hline 1.000 \quad (\text{低位}) \end{array}$$


# 进制间的转换

## ▶ 十进制数转换成二进制数

### ▶ 将 $(19.125)_{10}$ 转换成二进制数

▶  $(19)_{10} = (10011)_2$

▶  $(0.125)_{10} = (0.001)_2$

▶  $\therefore (19.125)_{10} = (10011.001)_2$

### ▶ Tip: 并非所有的十进制小数都能用有限位的二进制小数来表示!

### ▶ 将 $(0.63)_{10}$ 转换为二进制

▶ 因为, 小数部分乘以2会无限循环下去

▶  $\therefore (0.63)_{10} = (0.1010)_2$

$$\begin{array}{r} 0.63 \\ \times \quad 2 \\ \hline 1.26 \\ \times \quad 2 \\ \hline 0.52 \\ \times \quad 2 \\ \hline 1.04 \\ \times \quad 2 \\ \hline 0.08 \end{array}$$

# 进制间的转换

## ▶ 二进制与八进制之间的转换

$$\text{▶ } (011\ 110\ 111.\ 100\ 010\ 101)_2 = (367.425)_8$$

$$3\quad 6\quad 7\quad .\quad 4\quad 2\quad 5$$

$$\begin{aligned}\text{▶ } (16.327)_8 &= (001\ 110.\ 011\ 010\ 111)_2 \\ &= (1110.011010111)_2\end{aligned}$$

# 进制间的转换

## ▶ 二进制与十六进制之间的转换

$$\text{▶ } (1110\ 0101\ 1010 . 1011\ 1001)_2 = (E5A.B9)_{16}$$

E 5 A . B 9

$$\text{▶ } (4C.2E)_{16} = (0100\ 1100 . 0010\ 1110)_2$$

4 C . 2 E

$$= (1001100.0010111)_2$$

# 机器数与真值

- ▶ 由于计算机只能直接识别和处理用0,1两种状态表示的二进制形式的数据,所以在计算机中无法按日常的书写习惯.
- ▶ 用正负号加绝对值来表示数值,而与数字一样,需要用二进制代码0和1来表示正、负号.
- ▶ 最高位作为符号位,0表示正,1表示负.
  - ▶ 00100100 = + 36
  - ▶ 10100100 = - 36
- ▶ 而与机器数对应的用正,负符号加绝对值来表示的实际数值称为真值.

# 数据的原码表示

- ▶ 为了解决表示一个数字的符号的问题, 首先的处理办法应该是分配一个符号位来表示这个符号: 设置这个位(通常为最高有效位)为0表示一个正数, 为1表示一个负数. 数字中的其它位指示数值(或者绝对值). 因此一个字节只有7位(除去符号位), 数值的范围从0000000(0)到1111111(127).
- ▶ 这样当你增加一个符号位(第八位)后,
- ▶ **原码的表示范围:  $[-127_{10}, +127_{10}]$ .**
- ▶ Example:
  - ▶  $00100100 = +36$        $(+0)_{\text{原}} = 00000000$
  - ▶  $10100100 = -36$        $(-0)_{\text{原}} = 10000000$

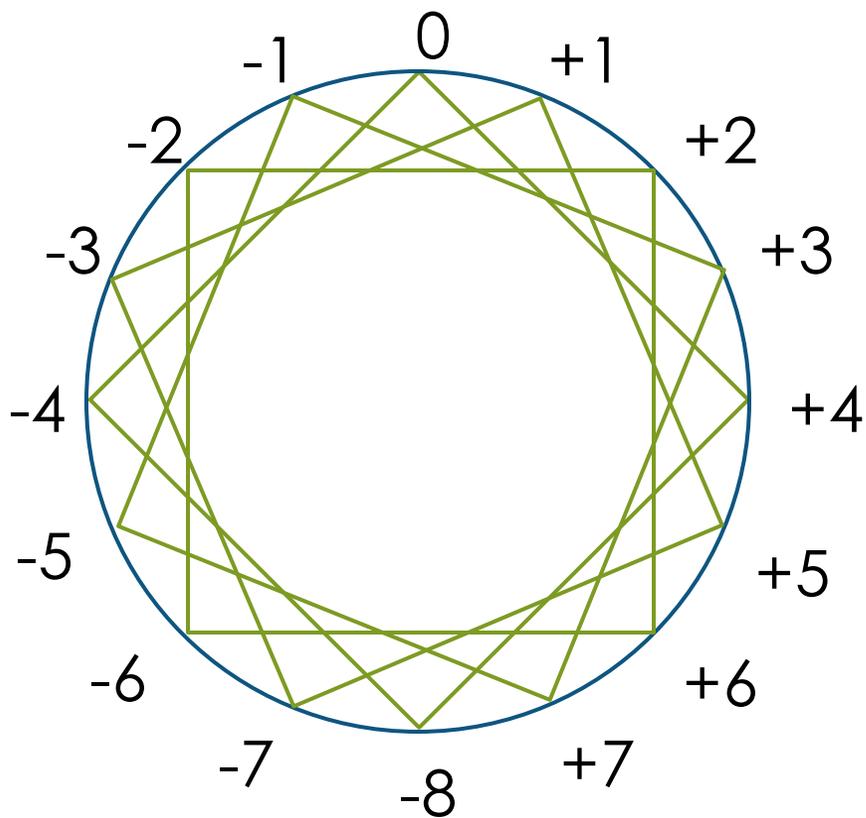
# 数据的反码表示

- ▶ 反码系统也可以用于表示负数. 一个负二进制数的反码形式为符号位不变, 其余位在原数基础上按位取反.
- ▶ 特别地: 规定正数的原码和反码相等.
  - ▶  $(+36)_{\text{原}} = (+36)_{\text{反}} = 00100100$
- ▶ **反码的表示范围:  $[-127_{10}, +127_{10}]$ .**
- ▶ Example:
  - ▶  $(-36)_{\text{原}} = 10100100$      $(+0)_{\text{反}} = 00000000$
  - ▶  $(-36)_{\text{反}} = 11011011$      $(-0)_{\text{反}} = 11111111$

# 数据的补码表示

- ▶ 补码系统中回避了0有多种表示的问题以及循环进位的需要. 一个负二进制的补码形式在反码基础上加1.
- ▶ 特别地:规定正数的原码、反码和补码相等.
  - ▶  $(+36)_{\text{原}} = (+36)_{\text{反}} = (+36)_{\text{补}} = 00100100$
- ▶ 补码的表示范围:  $[-128_{10}, +127_{10}]$ .
- ▶ Example:
  - ▶  $(-36)_{\text{原}} = 10100100$
  - ▶  $(-36)_{\text{反}} = 11011011$
  - ▶  $\quad\quad\quad + \quad\quad\quad 1$
  - ▶  $(-36)_{\text{补}} = 11011100$

# 数据的补码表示



▶ 假定现在是北京时间6点整,有一只手表却是8点整.校准方法有两种:

- ▶ 倒拨2小时
- ▶ 正拨10小时

▶ 一般性推广:

- ▶  $M \Rightarrow$  模
- ▶  $A - B = A + (-B + M) = A + (-B)_{\text{补}}$

# 定点数据和浮点数据

- ▶ 定点数据的两种形式:

- ▶ 计算机中并不表示出小数点, 而是人为地作出规定.

- ▶ 如规定小数点隐含在有效数字的最后, 则表示的是整数

- ▶ 如规定小数点隐含在有效数字的前面, 则表示的是纯小数

- ▶ 例如, 二进制数 $11011101$ , 可表示

- ▶ 整数:       -  $1011101$

- ▶ 纯小数: -  $0.1011101$

# 定点数据的表示范围

- ▶ 长度（包括符号位）为一个字节(8bit)的一个数值数据,可表示的整数的范围为:
  - ▶ 原码:  $-(2^7 - 1) \leq N \leq 2^7 - 1 = 127$
  - ▶ 反码:  $-(2^7 - 1) \leq N \leq 2^7 - 1 = 127$
  - ▶ 补码:  $-128 \leq N \leq 127$
- ▶ 有符号数据的正数范围
  - ▶ 16 Bit 二进制数  $-32768 \sim +32767$
  - ▶ 32 Bit 二进制数  $-2147483648 \sim +2147483647$
- ▶ 无符号数据的正数范围
  - ▶ 16 Bit 二进制数  $0 \sim 65535$
  - ▶ 32 Bit 二进制数  $0 \sim 2^{32} - 1 = 4294967295$

# 数据的浮点表示

- ▶ 在一般数据的浮点表示法中, 一个数可表示成 :

$$N = M \cdot R^E$$

E: 阶码, R: 基数, M: 尾数

- ▶ 计算机内部, 以浮点形式表示的数的基数通常隐含为2( $R = 2$ ), 在给定的字长中, 给出阶码和尾数的位数, 即可表示一个浮点数.

Single precision



Double precision



# 数据的浮点表示

- ▶ 写出二进制数 $-101.1101(B)$ 的浮点数形式, 设阶码取4位补码, 尾数是8位原码.

- ▶  $-101.1101 = -0.1011101 \times 2^{+3}$



# 计算机中的信息编码

## ▶ 西文信息在计算机内的表示

- ▶ ASCII(American Standard Code for Information Interchange, 美国信息交换标准代码)是基于拉丁字母的一套电脑编码系统. 它主要用于显示现代英语.
- ▶ ASCII码是7位编码, 从0000000到1111111共128种编码, 可用来表示128个不同的字符编码, 即可用来表示128个不同的字符.
  - ▶ 用一个字节存放一个ASCII码, 高位为0.
  - ▶ 其中:可打印字符95个, 在键盘上有相应键位, 如字母、数字等;
  - ▶ 控制字符33个, 在传输、打印或显示输出时起控制作用.

# 计算机中的信息编码

## ▶ 中文信息在计算机内的表示

- ▶ 汉字采用《信息交换用汉字编码字符集》(GB2312 - 80). 该标准称为“国标码” .
- ▶ 国标字符集有6763个常用汉字
- ▶ 国标字符集由三部分组成：
  - ▶ 字母、数字和各种符号, 共687个;
  - ▶ 一级常用汉字, 共3755个, 按汉语拼音排序;
  - ▶ 二级常用汉字, 共3008个, 按偏旁部首排序.

# 计算机中的信息编码

- ▶ 中文信息在计算机内的表示
  - ▶ 国标码是一个二维代码表, 有94行、94列
  - ▶ 汉字在代码表中的位置用它所处的行号、列号表示 :
  - ▶ 行号→区号      列号→位号
- ▶ 信息交换用汉字编码字符集
  - ▶ 汉字区位码: 汉字在代码表中的区号、位号
  - ▶ 汉字国标码: 区号、位号各增加32
  - ▶ 汉字机内码: 每个汉字的国标码可用14个二进制位表示, 在计算机中占两个字节, 每个字节高位置 “1”

# 计算机中的信息编码

## ▶ 区位码、 国标码、 机内码之间的转换

### ▶ 假设某个汉字

▶ 国标码G1 - G2, 区位码Q1 - Q2, 机内码 J1 - J2.

### ▶ 则有:

▶  $Q1 + 32 = G1$      $Q2 + 32 = G2$

(\*)为了区分ASCII码中的控制字符

▶  $G1 + 128 = J1$      $G2 + 128 = J2$

(\*)为了区分中文和英文

# 计算机中的信息编码

## ▶ 其他汉字字符集与编码

- ▶ GBK: 16位字符集; 收录有6763个简体汉字, 682个符号, 共7445个字符
- ▶ GB2312: 16位字符集; 收录有21003个汉字, 883个符号, 共21886个字符
- ▶ GB18030: 32位字符集; 收录了27484个汉字, 同时收录了藏文、蒙文、维吾尔文等主要的少数民族文字
- ▶ Unicode: 32位字符集; 包含了超过十万个字符. 它为每种语言中的每个字符设定了统一并且唯一的二进制编码
- ▶ UTF-8: 变长字符集, 使用8~32位存储一个字符; UTF-8就是在互联网上使用最广的一种Unicode的实现方式

# 汉字的输入与输出

## ▶ 汉字的输入

### ▶ 汉字输入方法大体可分为: 区位码、音码、形码、音形码

- ▶ 区位码: 优点是无重码或重码率低, 缺点是难于记忆;
- ▶ 音码: 优点是大多数人都易于掌握, 但同音字多, 重码率高, 影响输入的速度;
- ▶ 形码: 根据汉字的字型进行编码, 编码的规则较多, 难于记忆, 必须经过训练才能较好地掌握;重码率低;
- ▶ 音形码: 将音码和形码结合起来, 输入汉字, 减少重码率, 提高汉字输入速度



# 汉字的输入与输出

## ▶ 汉字的输出

### ▶ 矢量字体(Vector Font)

- ▶ 矢量字体是与点阵字体相对应的一种字体. 矢量字体的每个字形都是通过数学方程来描述的, 一个字形上分割出若干个关键点, 相邻关键点之间由一条光滑曲线连接, 这条曲线可以由有限个参数来唯一确定. 矢量字的好处是字体可以无级缩放而不会产生变形. 向量字体在显示或者打印出来之前需要进行栅格化. 目前主流的矢量字体格式有3种: Type1, TrueType和OpenType, 这三种格式都是平台无关的.



# Thank You!

谢浩哲

[cshzxie@gmail.com](mailto:cshzxie@gmail.com)