

前言

作为一个程序员，需要了解存储器层次结构，因为它对性能有着巨大的影响

计算机中基本而持久的思想

存储技术分类

随机访问存储器

随机存储器的分类：静态RAM和动态RAM

非易失性存储器

磁盘存储

磁盘构造

磁盘容量

磁盘操作

固态硬盘

现有常见存储器的分类

发展趋势

存储器的层次结构

存储器层次结构中的缓存

高速缓存存储器

前言

作为一个程序员，需要了解存储器层次结构，因为它对性能有着巨大的影响

如果需要的数据存储在cpu的寄存器中，在0个周期内，cpu就可以访问到他们

如果要的数据存储在高速缓存中，需要4 - 75个周期，cpu才能访问到他们

如果要的数据存储在主存中，需要上百个周期，cpu才能访问到他们

如果存储在硬盘上，需要大约几千万个周期

计算机中基本而持久的思想

数据存储在层次结构越高的地方，cpu就能够更快的访问他们

存储技术分类

随机访问存储器

随机存储器的分类：静态RAM和动态RAM

随机访问存储器（Random-Access Memory，RAM）分为两类：静态的和动态的

静态RAM（STAM）比动态RAM（DRAM）更快，但也贵得多。

静态RAM可以作为高速缓存寄存器，动态RAM可以用作主存以及图形系统的帧缓冲区

	每位晶体管数	相对访问时间	持续的？	敏感的？	相对花费	应用
SRAM	6	1×	是	否	1000×	高速缓存存储器
DRAM	1	10×	否	是	1×	主存，帧缓冲区

非易失性存储器

如果断电，DRAM 和 SRAM 都会丢失信息，从这个方面来说，它们是易失的（volatile）

如果断电，还能保存信息，这种就是非易失性存储器

只读存储器ROM(Read Only Memory)，每个单元都有一种熔丝，只能用高电流熔断一次，读取的时候电流比较低，可以反复读取。写的时候电流比较高，写完之后熔丝断裂，之后只能读取，无法写入

可擦写存储器，可以反复读写，但是次数有限。不同的物理材料 可以擦写次数也不一样。

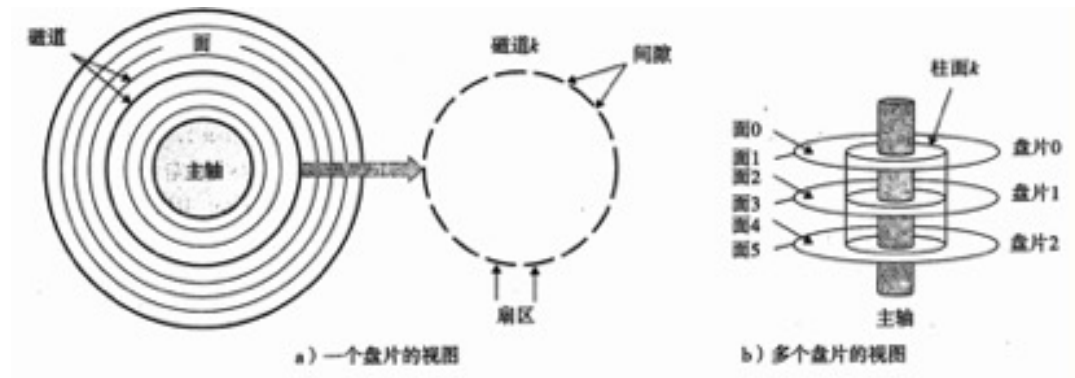
闪存是一种非易失性存储器，它无处不在，为大量电子设备提供了快速而持久的非易失性存储。手机、pad、数码相机等

磁盘存储

磁盘是用来保存大量数据的存储设备，能比基于RAM的存储器存更多的数据，但是读取数据的速度却远远低于基于RAM的存储器，磁盘速度比DRAM慢了10万倍，磁盘速度比SRAM慢了100万倍

磁盘构造

磁盘是由盘片构成的。每个盘片上是由一组成为磁道的同心圆组成的。每个磁道被划分为一组扇区。扇区之间存在间隙



磁盘容量

磁盘容量是由以下因素决定的：记录密度、磁道密度、面密度

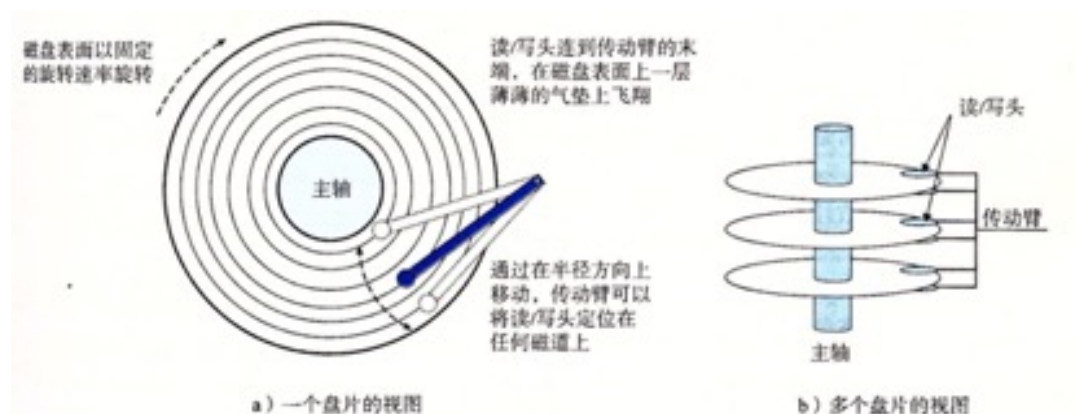
磁盘操作

传动臂通过寻道操作将读写头定位到相应的磁道上。磁盘以扇区大小的块来访问数据，对于扇区的访问时间有三个部分：

寻道时间：移动传动臂所需的时间

旋转时间：到达期望磁道后，驱动器等待目标扇区的第一个位旋转到读/写头下的时间

传送时间：第一个位位于读/写头下时，驱动器开始读或者写该扇区内容的时间



估计访问下面这个磁盘上一个扇区的访问时间（单位：ms）

参数	值
旋转速率	15000RPM
T(avg seek)	8ms
每条磁道的平均扇区数	500

$T(\text{平均寻道时间}) = 8\text{ms}$

$T(\text{平均旋转时间}) = 1/2 * 1 / 15000(\text{r/m}) * 60(\text{s/m}) * 1000(\text{ms/s}) * 1(\text{r}) = 2\text{ms}$

$T(\text{平均传送时间}) = 1 / 15000(\text{r/m}) * 60(\text{s/m}) * 1000(\text{ms/s}) * 1(\text{r}) / 500 = 0.008\text{ms}$

所以 $T(\text{访问时间}) = T(\text{平均寻道时间}) + T(\text{平均旋转时间}) + T(\text{平均传送时间}) = 10.008\text{ms}$

固态硬盘

固态硬盘（Solid State Disk,SSD），基于闪存的存储技术。一个SSD封装由一个或多个闪存芯片和闪存翻译层组成

由半导体存储器组成，没有移动的部件，所以随机访问时间比旋转磁盘要快，能耗更低，也更结实

现有常见存储器的分类



易失性存储器就是和非易失性存储器的唯一区别在于前者掉电数据会被清除。

1、 易失性存储器的代表就是RAM，RAM又分DRAM（动态随机存储器）和SRAM（静态随机存储器），他们之前不同在于生产工艺的不同，SRAM保存数据是靠晶体管锁存的，DRAM保存数据靠电容充电来维持。SRAM的工艺复杂，生产成本低，所以贵，容量比较大的RAM我们都选用的是DRAM。而且SRAM速度较快。

2、 RAM既然是存储器就要传输数据，传输数据就是通信。通信又分同步通信和异步通信。前面我们所说的DRAM和SRAM都是异步通信的，速率没有SDRAM和SSRAM快。所以现在大容量RAM存储器是选用SDRAM的。S(Synchronous同步)

3、 现在电脑里面用的内存条就是RAM。我的电脑台式机用的是DDR3SDRAM，我的手机用的是DDR4 SDRAM，我的嵌入式开发板用的是DDR2 SDRAM。那么DDR SDRAM和SDRAM的区别在于DDR（double data rate）双倍

速率。SDRAM只在时钟的上升沿表示一个数据，而DDR SDRAM能在上升沿和下降沿都表示一个数据。DDR也一步步经过改良出现了一代、二代、三代，现在也有四代。

4、 Cache常见于CPU中，cache实质属于SRAM，所以是造价高，但是速度快，比DRAM快，在电脑中体现在cache（sram）比内存（dram）快，所以cache作为CPU和内存之间通信的桥梁。

5、 那么cache是怎么加快CPU和内存的通信的呢？先了解cache是要解决CPU和内存之间的什么矛盾。矛盾在于：CPU读取速度快，而内存给的速度慢，这样CPU要接一会数据等一会，浪费了CPU处理时间。我们把CPU常读取的内存的数据放到cache中，CPU读取cache很快，这样CPU就免了等待时间，CPU的处理速度就提高了。还有一个问题就是cache怎么知道哪些是CPU的常用数据？其实cache存储的是PU刚用过或循环使用的一部分数据，就是做一些数据的缓存。所以cache又叫缓存。

6、 非易失性存储器常见的有ROM，FLASH，光盘，软盘，机械硬盘。他们作用相同，只是实现工艺不一样。

7、 光盘、软盘和机械硬盘都很好理解，不做解释。

8、 ROM(Read Only Memory)在以前就是只读存储器，就是说这种存储器只能读取它里面的数据无法向里面写数据。实际是以前向存储器写数据不容易，所以这种存储器就是厂家造好了写入数据，后面不能再次修改。现在技术成熟了，ROM也可以写数据，但是名字保留了下来。

9、 ROM分为MASK ROM、OTPROM、EPROM、EEPROM。MASK ROM是掩膜ROM这种ROM是一旦厂家生产出来，使用者无法再更改里面的数据。

OTPROM(One Time Programmable ROM)一次可变成存储器，出厂后用户只能写一次数据，然后再也不能修改了，一般做存储密钥。EPROM(Erasable Programmable ROM)这种存储器就可以多次擦除然后多次写入了。但是要在特定环境紫外线下擦除，所以这种存储器也不方便写入。EEPROM(Electrically Erasable Programmable ROM)电可擦除ROM，现在使用的比较多因为只要有电就可擦除数据，就可以写入数据。

10、 FLASH是一种可以写入和读取的存储器，叫闪存，FLASH也叫FLASH ROM，有人把FLASH当做ROM。FLASH和EEPROM相比，FLASH的存储容量大。FLASH的速度比现在的机械硬盘速度快，现在的U盘和SSD固态硬盘都是Nandflash。FLASH又分为Norflash和Nandflash

发展趋势

SRAM

度量标准	1985	1990	1995	2000	2005	2010	2015	2015:1985
美元/MB	2900	320	256	100	75	60	25	116
访问时间 (ns)	150	35	15	3	2	1.5	1.3	115

DRAM

度量标准	1985	1990	1995	2000	2005	2010	2015	2015:1985
美元/MB	880	100	30	1	0.1	0.06	0.02	44 000
访问时间 (ns)	200	100	70	60	50	40	20	10
典型的大小 (MB)	0.256	4	16	64	2000	8000	16 000	62 500

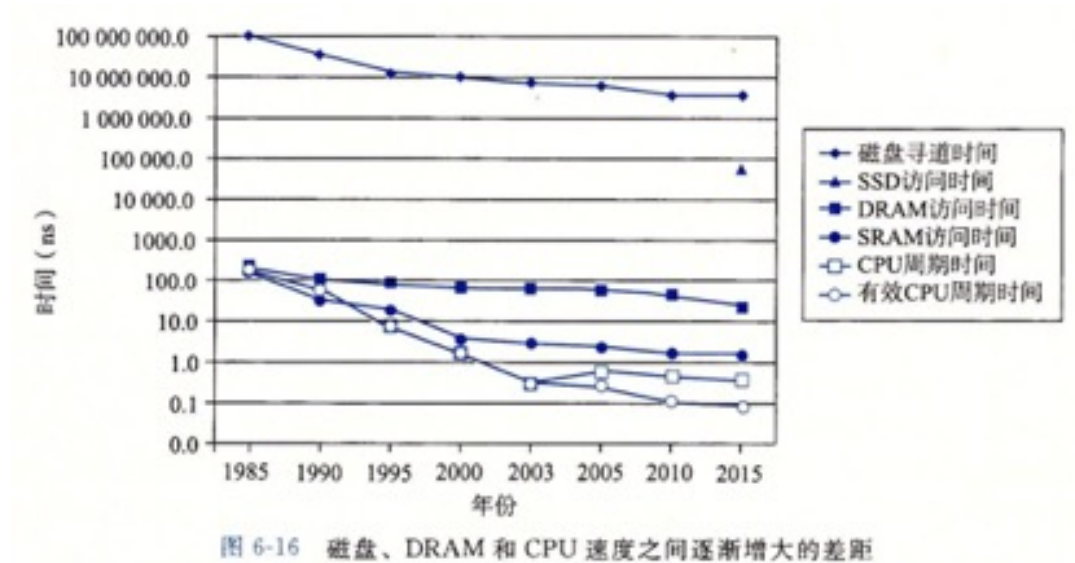
机械磁盘

度量标准	1985	1990	1995	2000	2005	2010	2015	2015:1985
美元/GB	100 000	8000	300	10	5	0.3	0.03	3 333 333
最小寻道时间 (ms)	75	28	10	8	5	3	3	25
典型的大小 (GB)	0.01	0.16	1	20	160	1500	3000	300 000

CPU

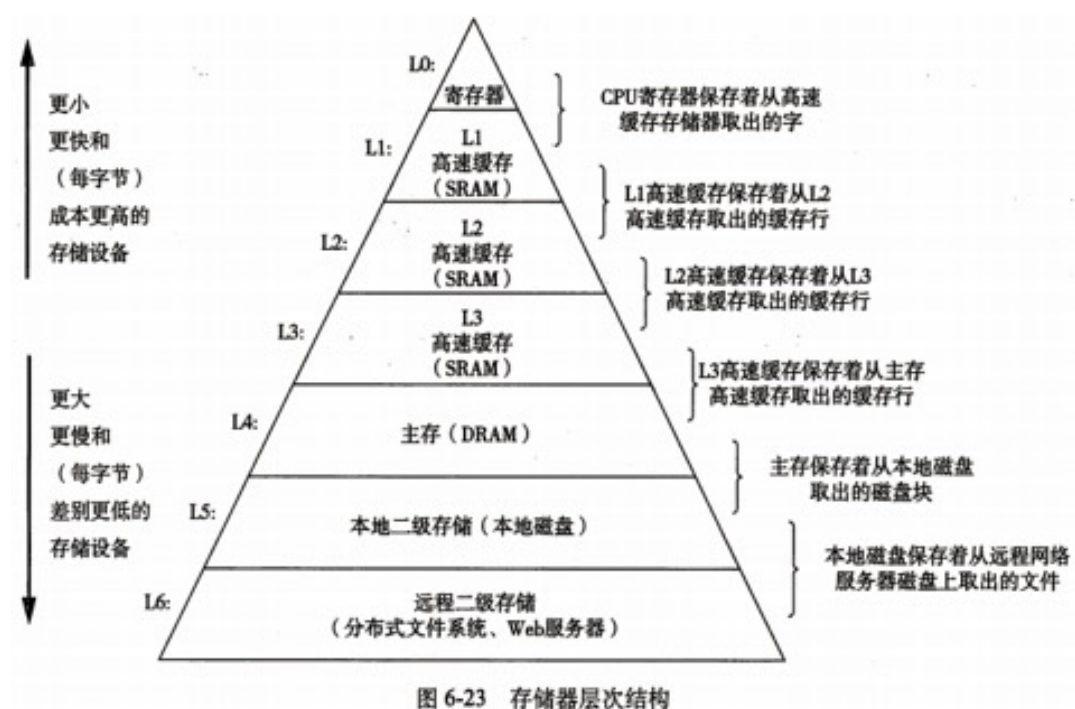
度量标准	1985	1990	1995	2000	2003	2005	2010	2015	2015:1985
Intel CPU	80 286	80 386	Pent.	P-III	Pent.4	Core 2	Core i7 (n)	Core i7 (h)	—
时钟频率 (MHz)	6	20	150	600	3300	2000	2500	3000	500
时钟周期 (ns)	166	50	6	1.6	0.3	0.5	0.4	0.33	500
核数	1	1	1	1	1	2	4	4	4
有效周期时间 (ns)	166	50	6	1.6	0.30	0.25	0.10	0.08	2075

磁盘、DRAM和CPU速度之间的逐渐增大的差距



存储器的层次结构

所有的现代计算机系统都使用存储器结构层次来使得软件和硬件互相补充



一般而言，从高层往底层走，存储设备变得更慢、更便宜和更大

存储器层次结构中的缓存

一般而言，高速缓存是一个小而快速的存储设备，它作为存储在更大也更慢的存储设备中的数据对象的缓冲区域。使用高速缓存的过程称为“缓存”

概括来说，基于缓存的存储器层次结构行之有效，是因为较慢的存储设备比较快的存储设备更便宜，还因为程序往往展现局部性：利用时间局部性和利用空间局部性

高速缓存存储器

早期计算机系统的存储结构只有三层：CPU寄存器、DRAM主存储器和磁盘存储。不过，由于CPU和主存之间逐渐增大的距离，系统设计者被迫在CPU寄存器文件和主存之间插入了一个小的SRAM高速缓存存储器，称为L1高速缓存。随着CPU和主存之间的性能差距不断增大，系统设计者在L1高速缓存更大的高速缓存，称为L2高速缓存。有些现代计算机还包括一个更大的高速缓存，称为L3缓存