

存储系统存在的问题

双端口RAM、多模块存储器提高存储器的工作速度

优化后速度与CPU差距
依然很大

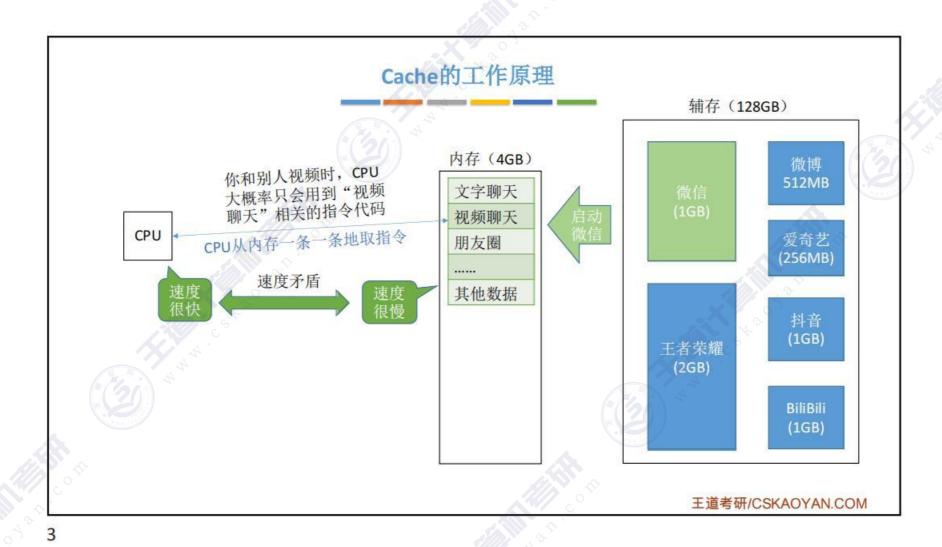
更高速的存储单元设计

存储器价格↑容量↓

存储体系的改善

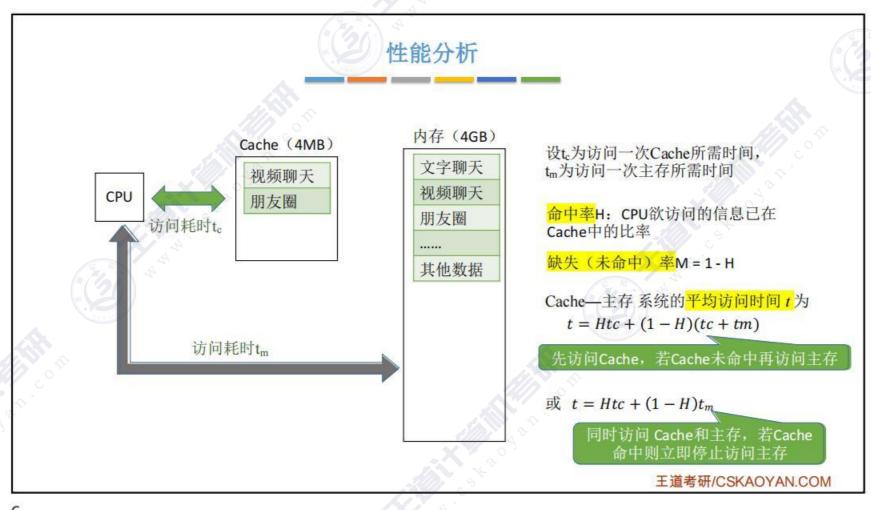
"Cache-主存"层次

程序访问的局部性原理



Cache的工作原理 辅存 (128GB) 注:实际上,Cache 被集成在CPU内部Cache用SRAM实现,速度快,成本高 内存 (4GB) Cache (4MB) 微博 512MB 文字聊天 视频聊天 视频聊天 CPU 爱奇艺 朋友圈 (256MB) 其他数据 Cache比内 存快60倍 速度矛盾被缓和 王者荣耀 (2GB) AIDA64 Cache & Memory Benchmark 三届内存京东企业师 988.09 GB/s 497.62 GB/s 992.73 GB/s 王道考研/CSKAOYAN.COM

局部性原理 Eg: 数组元素、顺序执 行的指令代码 程序A: 0x0FC 空间局部性: 在最近的未来要用 int sumarrayrows (int a[M][N]) 0x100 11 2 到的信息(指令和数据),很可能 0x104 3 int i, j, sum = 0; 与现在正在使用的信息在存储空 0x108 for (i = 0; i < M; i++)指 4 for内/外循环 间上是邻近的 0x10C for (j = 0; j < N; j++)5 0x110 6 sum += a[i][j]; 时间局部性:在最近的未来要用 0x114 136 7 return sum; 到的信息,很可能是现在正在使 用的信息 a[0][0] 0x400 程序B: a[0][1] 0x404 0x408 int sumarraycols(int a[M][N]) a[0][2047] 0x40C 基于局部性原理, 不难想到, 可 0x410 a[1][0] 以把CPU目前访问的地址"周围" int i, j, sum = 0; 0x414 a[1][1] 的部分数据放到Cache中 for (j = 0; j < N; j++)for (i = 0; i < M; i++)sum += a[i][j]; 程序B按"列优先"访问二维数 0x7A4 return sum; 组,空间局部性更差 指令和数据在内存中的存储 王道考研/CSKAOYAN.COM



性能分析

【例3-2】 假设Cache的速度是主存的5倍,且Cache的命中率为95%,则采用Cache后,存储器性能提高多少(设Cache和主存同时被访问,若Cache命中则中断访问主存)?

设Cache的存取周期为t,则主存的存取周期为5t

若Cache和主存同时访问,命中时访问时间为t,未命中时访问时间为5t 平均访问时间为 $0.95 \times t + 0.05 \times 5t = 1.2t$

故性能为原来的 <u>5t</u> ≈4.17倍

<mark>若先访问Cache再访问主存</mark>,命中时访问时间为t,未命中时访问时间为 t+5t 平均访问时间为 T_a = $0.95 \times t + 0.05 \times 6t = 1.25t$

故性能为原来的 $\frac{5t}{1.25t}$ =4倍

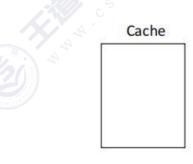
王道考研/CSKAOYAN.COM

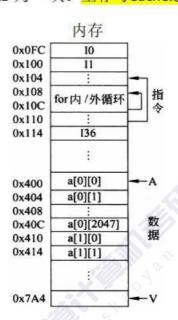
7

有待解决的问题

基于局部性原理,不难想到,可以把CPU目前访问的地址"周围"的部分数据放到Cache中。如何界定"周围"?

将主存的存储空间"分块",如:每1KB为一块。主存与Cache之间以"块"为单位进行数据交换



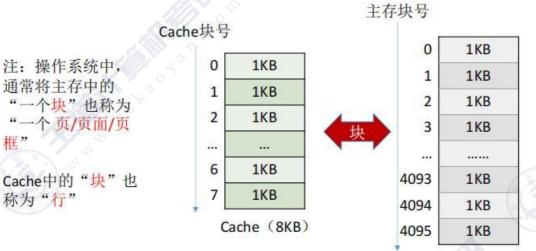


王道考研/CSKAOYAN.COM

有待解决的问题

基于局部性原理,不难想到,可以把CPU目前访问的地址"周围"的部分数据放到Cache中。如何界定 "周围"?

将主存的存储空间"分块",如:每1KB为一块。主存与Cache之间以"块"为单位进行数据交换



主存的地址共22位:

块号	块内地址
12位	10位

 $4M=2^{22}$, $1K=2^{10}$ 整个主存被分为 212 = 4096 块

王道考研/CSKAOYAN.COM

注:操作系统中,

通常将主存中的

称为"行"

"一个块"也称为

"一个页/页面/页

有待解决的问题

主存 (4MB)

主存块号 Cache块号 0 1KB 0 1KB 1KB 1 1 1KB 2 1KB 2 1KB 3 1KB 6 1KB 4093 1KB 1KB 4094 1KB Cache (8KB) 4095 1KB 主存 (4MB)

注意:每次被访问的主存块, 一定会被立即调入Cache

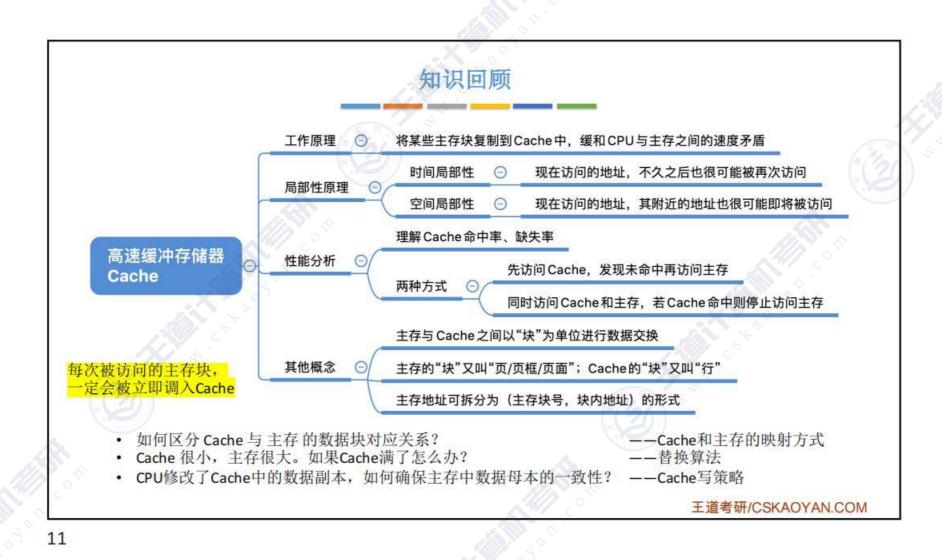
主存的地址共22位:

块号	块内地址
12位	10位

4M=2²², 1K=2¹⁰ 整个主存被分为 212 = 4096 块

- 如何区分 Cache 与 主存 的数据块对应关系?Cache 很小, 主存很大。如果Cache满了怎么办?
- CPU修改了Cache中的数据副本,如何确保主存中数据母本的一致性?
- ——Cache和主存的映射方式
- -替换算法
- ——Cache写策略

王道考研/CSKAOYAN.COM



你还可以在这里找到我们

快速获取第一手计算机考研信息&资料



购买2024考研全程班/领学班/定向班 可扫码加微信咨询

- 微博: @王道计算机考研教育
- B站: @王道计算机教育
- 小红书: @王道计算机考研
- 知 知乎: @王道计算机考研
- 抖音: @王道计算机考研
- 淘宝:@王道论坛书店