УДК 004.252

Artem Klokov[1],Olha Kalantaieva[2]
[1]student of groupCST-118 ZNTU
[2]teacher ZNTU

MEMORY HIERARCHY

In computer architecture, the memory hierarchy separates computer storage into a hierarchy based on response time. Since response time, complexity, and capacity are related, the levels may also be distinguished by their performance and controlling technologies. Memory hierarchy affects performance in computer architectural design, algorithm predictions, and lower level programming constructs involving locality of reference.

Designing for high performance requires considering the restrictions of the memory hierarchy, i.e. the size and capabilities of each component. Each of the various components can be viewed as part of a hierarchy of memories (m_1,m_2,...,m_n) in which each member $m_{(i)}$ is typically smaller and faster than the next highest member $_{(i+1)}$ of the hierarchy. To limit waiting by higher levels, a lower level will respond by filling a buffer and then signaling for activating the transfer.

There are four major storage levels:1. Internal – Processor registers and cache; 2. Main – the system RAM and controller cards; 3. On-line mass storage – Secondary storage; 4. Off-line bulk storage – Tertiary and Off-line storage.

This is a general memory hierarchy structuring. Many other structures are useful. For example, a paging algorithm may be considered as a level for virtual memory when designing a computer architecture, and one can include a level of nearline storage between online and offline storage.

Properties of the technologies in the memory hierarchy

Adding complexity slows down the memory hierarchy.

CMOx memory technology stretches the Flash space in the memory hierarchy. One of the main ways to increase system performance is minimising how far down the memory hierarchy one has to go to manipulate data. Latency and bandwidth are two metrics associated with caches and. Neither of them is uniform, but is specific to a particular component of the memory hierarchy. Predicting where in the memory hierarchy the data resides is difficult.

Examples

The number of levels in the memory hierarchy and the performance at each level has increased over time. The type of memory or storage components also change historically. For example, the memory hierarchy of an Intel Haswell Mobile processor circa 2013 is:

1.Processor registers – the fastest possible access (usually 1 CPU cycle). A few thousand bytes in size; 2.Cache; 3.Main memory – Gigabytes in size. Best access speed is around 10 GB/second; 4.Disk storage (Secondary storage) – Terabytes in size. Best access speed is from a consumer solid state drive is about 2000 MB/second; 5.Nearline storage (Tertiary storage) – Up to exabytes in size. Best access speed is about 160 MB/second; 6.Offline storage.

The lower levels of the hierarchy – from disks downwards – are also known as tiered storage is:

1. Online storage is immediately available for I/O;

2.Nearline storage is not immediately available, but can be made online quickly without human intervention;

3.Offline storage is not immediately available, and requires some human intervention to bring online.