CMP 301 19/01/2024

Memory system design 1

When talking about memory we consider the hierarchy, it is categorized into parameters which include:

* Access type: how you access the documents
* Capacity: it is measures in bytes
* Cycle time: this is the time for a cycle to complete, it is the time from one operation starting to the next operation start
* Latency: time between the request and the response
* Bandwidth: number of bits per second that can be accessed
* cost

types of access:

Random access: time to access any random document or information

CPU Registers -> cache -> Mani memory -> secondary -> tertiary storage ->

Latency and bandwidth decreases while the speed and cost per bit increases

Cache is used to store small bits of data so that the cpu can use it later without having to go to the ram or hard drive.

A principle called locality is used by the cpu to access information

Type

* spatial locality basically uses referencing by time the later the reference the easier the access.

Cache memory

When a processor makes a request it is sent to the cache if found by the cache it is a hit if not found it is called a miss

The cache hit ratio is the probability of a requested information to be found by the cache the miss ratio is 1 –HR. the advantage of transferring a block from the main memory to a cache is that if the requested information or its neighbor will be requested next time, it can be easily accessed

Formula for calculating

J=cache

I=memory

N=total no of

The remainder is the memory location depending on the task view basically base

View types (direct mapping)

* word view log2 B size of block in words
* block view log2 N size of cache in words
* tag view = log2 (M/N) main memory in block
* log2 (B \* M) bits in main memory address

note word + block + tag = total bits

(Fully associative mapping)

* word field = log2 B
* tag view = log2 M
* total = log2 (B\*M)

(set associative mapping)

* same as direct but N =S is the number of sets in the cache and s = N/B
* meaning the total number of bits divided by the number of each bit per set gives the total number of sets

Replacements techniques

* Random selection
* First in first out (longest goes out
* Least used block LRU