What is Cache memory ?

Cache memory is an extremely fast memory used in computers to compensate for the performance difference between main memory access time and processor logic.

Cache memory is placed between the CPU and the main memory.

The cache memory access time is 5 to 10 times faster than the main memory access time.

The cache stores programme segments that are currently being run in the CPU as well as data that is frequently used in current calculations.

It is possible to boost a computer’s performance rate by making programmes and data available at a faster rate.

L1 and L2 Cache Placement

L1 Cache : The cache memory placed inside the processor.

L2 Cache: The cache memory placed between L1 cache and main memory.

Cache replacement

Split Cache Organization

Split cache organization

Cache line

The cache memory is divided into several lines. These lines are known as “cache line”.

Where,

Size of cache line is same as the capacity of main memory block.

There are more blocks of main memory than number of cache lines in cache memory. Hence, cache maping techniques used to map any main memory block to one of the cache lines.

Now there is two situations:

1. Cache hit: If information already available in cache memory.
2. Cache miss: If information already not available in cache memory. Every cache miss required to access main memory, which requires extra time.

Q. In many computers the cache block size is in the range of 32 to 128 bytes. What would be the main advantages and disadvantages of making the size of cache blocks larger orsmaller?

Ans. Larger cache memory size

- Less case miss if most of the data in the block are actually used.
- Wasteful if much of the data are not used before the cache block is ejected from the cache.

Smaller size

- More cache mis, than extra time may required to search in main memory than copy to

the cache memory.