1. **Review Question No: 2**

Computer systems include multiple levels of cache, such as L1, L2, and L3, to make them faster and more effective. The CPU and main memory are both made closer to each level of cache than the other. This proximity makes the CPU speedier by reducing the time it takes to access data.

Larger, slower caches (like L2 and L3) are placed farther away from the CPU than smaller, faster caches (like L1), which make up a hierarchy of caches. The trade-off between speed and storage capacity is balanced by this configuration.

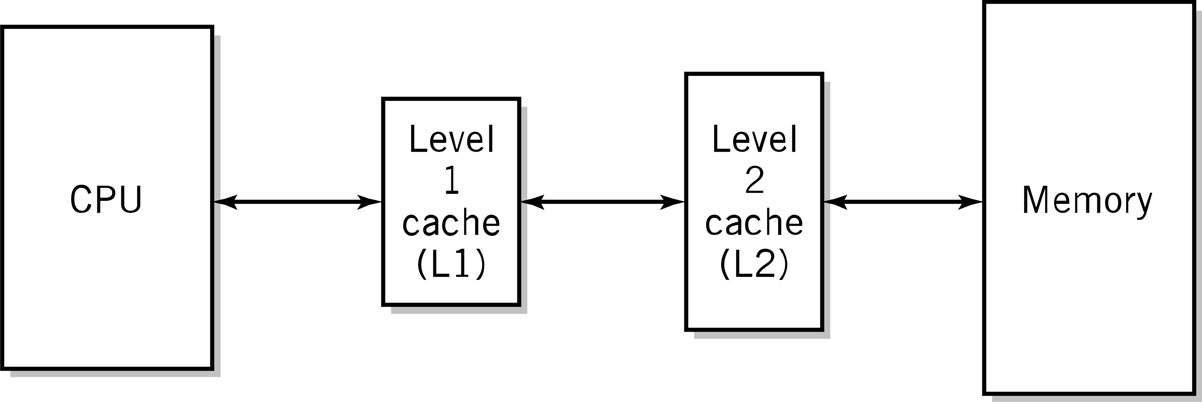
According to the principle that data that has recently been used is likely to be utilized again soon, caches keep track of frequently accessed information. The CPU may thus immediately access the data it requires without having to wait for the main memory, which is slower.

Caches aid in the coordination and synchronization of data between several CPUs in multiprocessor systems, ensuring that all cores have access to the most recent data.

Various cache levels also aid in controlling prices and energy usage. Although L1 caches are smaller and faster, they require more expensive and power-intensive technologies. The L2 and L3 caches, which are located further away, are bigger and slower but employ more economical and energy-efficient technology.

In general, having many cache levels aids in making computers speedier, decreasing delays when accessing data, enhancing performance, and effectively meeting the demands of contemporary CPUs.

The below Figure Shows an Example of Type of Cache Memories Available in the Computer



Figure

1. **Review Question No: 3**

To find a balance between speed and storage capacity, cache sizes must differ.

Consider the caches as data storage areas that retain frequently accessed information. L1 cache, or the closest cache to the CPU, is a tiny but quick cache. The CPU doesn't have to wait for very long because it can immediately give the data it needs.

However, it would be costly and impractical to store a lot of data in the quickest cache. There are thus more cache levels like the slower but larger L2 and L3. Although these caches have a larger storage capacity, accessing that data takes a little longer.

The computer system can place the most important and often used data in the smaller, quicker cache (L1) thanks to the varying cache sizes. The larger, slower caches (L2 and L3) can be used to store less often used information.

As a result, the various cache sizes allow for a balance between performance and storage capacity. While still allowing for more storage space for less crucial data, it makes sure that the most crucial and often accessible information is saved in the fastest cache.