Computer Architecture

NAME : Nabeel Hamid FAROOQI

Semester: Spring

ROLL NO : BSCS317-M/1-17/M022

SUBMITTED TO : Mrs.*Amina M*.Amin

DEPARTMENT : BSCS

Q1. What is the role of RAM while booting computer system ?

Ans) **The main function of RAM during booting is the main function of RAM any other time.**

RAM holds information that the processor is likely to need in pretty short order. In the instance of booting, this would be Operating System (OS) files from the hard drive. The hard drive is slow compared to RAM(especially older mechanical/magnetic hard drives, less true of solid-state drives, but still true). If the processor had to rely on hard-drive speeds for accessing everything, the system would grind to a halt. So, the processor loads things from the hard drive into RAM, where it can get to it quickly and use it. It takes a while to boot up, but once those files are in RAM, it can zip along at whatever speed it's capable of going.

If you don't have enough RAM to run the whole OS, less-frequently-used information can be cached back to the hard drive, but this will slow things down. This is why it's important to have enough RAM to run your OS and your programs, and some extra for doing temporary calculations and storage operations.

RAM is fast, but it's not permanent. As soon as the machine is shut down or loses power, anything stored in RAM is gone. This is what the hard drive is for. It can save things permanently.

So, a typical computer's power cycle looks roughly like this:

BIOS and Processor turn on.
Processor get info from BIOS on what hardware is available (RAM, hard drive, etc) and where it should look for boot files.

Boot files from the hard drive are loaded into memory (RAM) and executed by the processor.

Once the OS is loaded, programs are run (loaded into RAM, do some things) then closed (important things from RAM are written back to the hard drive for later use).

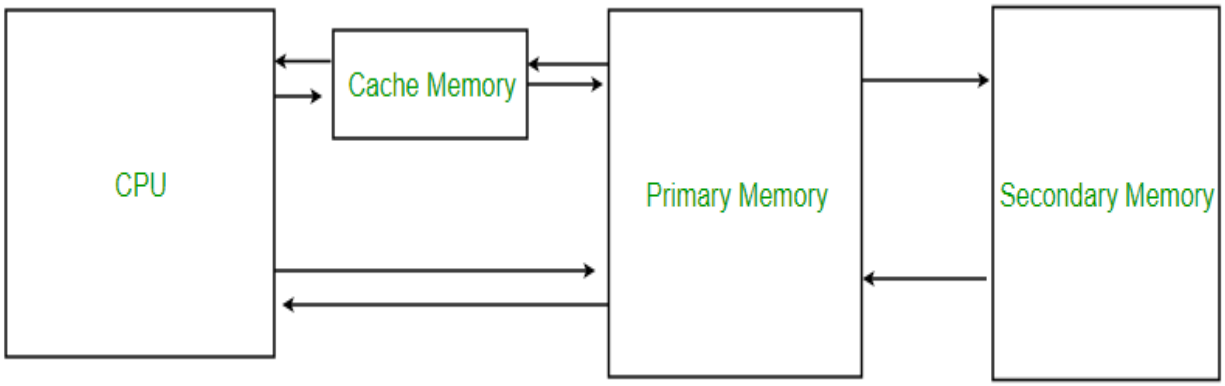
At shutdown, any data that must be saved for later is written to the hard drive. The processor completes save operations on files that had been loaded into RAM, then the machine shuts off. RAM becomes empty.

When a machine is powered off improperly, files that have not been saved fully to the hard drive can become corrupted and unreadable. This is because RAM was holding those changes and did not have time to write them back to the hard drive before it lost power.

In short, the purpose of RAM during booting is to take on the OS files so that the processor can effectively run the OS. And this is its function the rest of the time and for the other programs (OS is a program) the computer is running, too.

Q2. Describe the three types of cache?

Ans) **Cache Memory** is a special very high-speed memory. It is used to speed up and synchronizing with high-speed CPU. Cache memory is costlier than main memory or disk memory but economical than CPU registers. Cache memory is an extremely fast memory type that acts as a buffer between RAM and the CPU. It holds frequently requested data and instructions so that they are immediately available to the CPU when needed. Cache memory is used to reduce the average time to access data from the Main memory. The cache is a smaller and faster memory which stores copies of the data from frequently used main memory locations. There are various different independent caches in a CPU, which store instructions and data.



Levels of memory:

- **Level 1 or Register –**
  It is a type of memory in which data is stored and accepted that are immediately stored in CPU. Most commonly used register is accumulator, Program counter, address register etc.
- **Level 2 or Cache memory –**
  It is the fastest memory which has faster access time where data is temporarily stored for faster access.
- **Level 3 or Main Memory –**
  It is memory on which computer works currently. It is small in size and once power is off data no longer stays in this memory.
- **Level 4 or Secondary Memory –**
  It is external memory which is not as fast as main memory but data stays permanently in this memory.

Cache Performance:
When the processor needs to read or write a location in main memory, it first checks for a corresponding entry in the cache.

- If the processor finds that the memory location is in the cache, a **cache hit** has occurred and data is read from cache
- If the processor **does not** find the memory location in the cache, a **cache miss** has occurred. For a cache miss, the cache allocates a new entry and copies in data from main memory, then the request is fulfilled from the contents of the cache.

The performance of cache memory is frequently measured in terms of a quantity called **Hit ratio.**

```
Hit ratio = hit / (hit + miss) =  no. of hits/total accesses
```

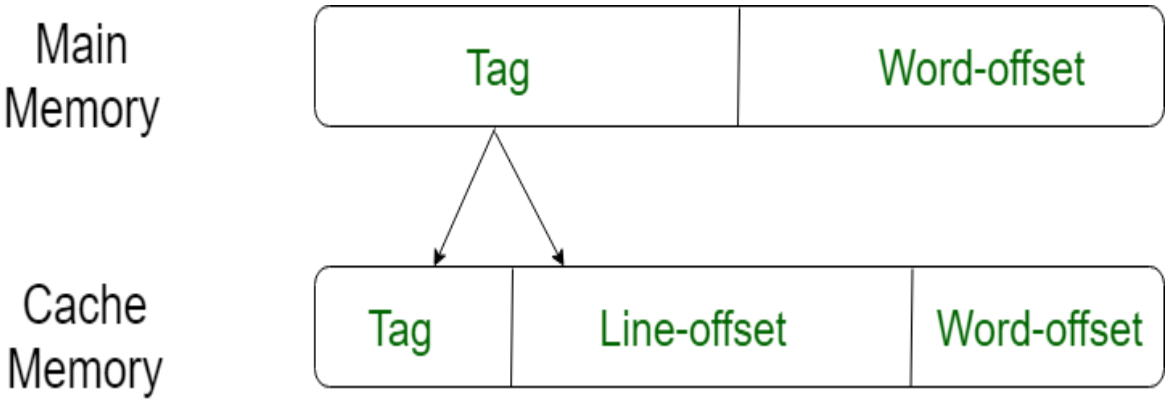
We can improve Cache performance using higher cache block size, higher associativity, reduce miss rate, reduce miss penalty, and reduce the time to hit in the cache.

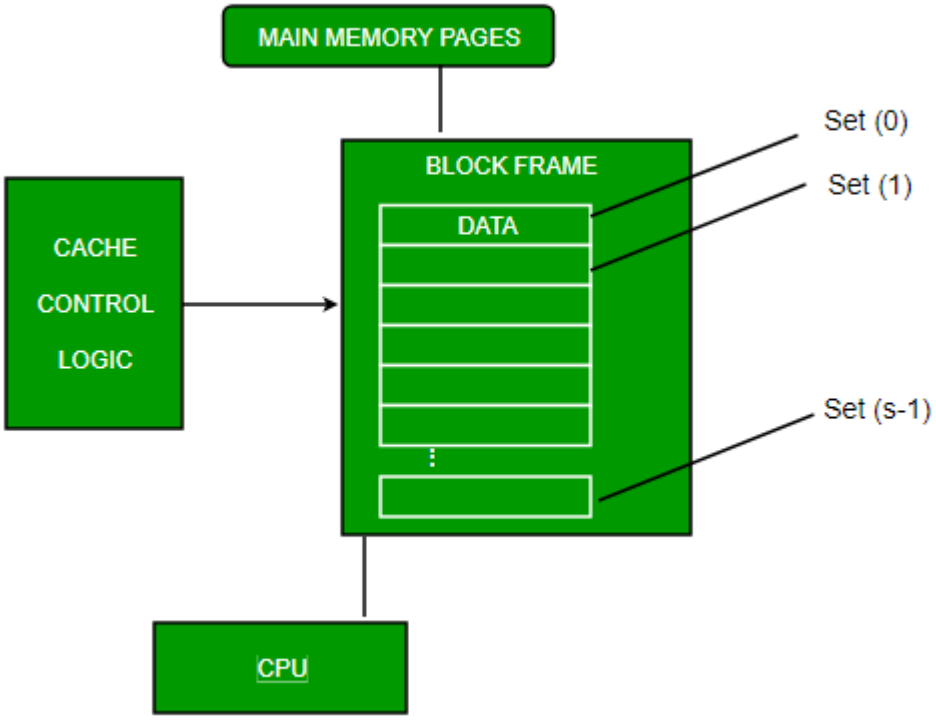
Cache Mapping:
There are three different types of mapping used for the purpose of cache memory which are as follows: Direct mapping, Associative mapping, and Set-Associative mapping. These are explained below.

1. **Direct Mapping –**
   The simplest technique, known as direct mapping, maps each block of main memory into only one possible cache line. or
   In Direct mapping, assigne each memory block to a specific line in the cache. If a line is previously taken up by a memory block when a new block needs to be loaded, the old block is trashed. An address space is split into two parts index field and a tag field. The cache is used to store the tag field whereas the rest is stored in the main memory. Direct mapping`s performance is directly proportional to the Hit ratio.

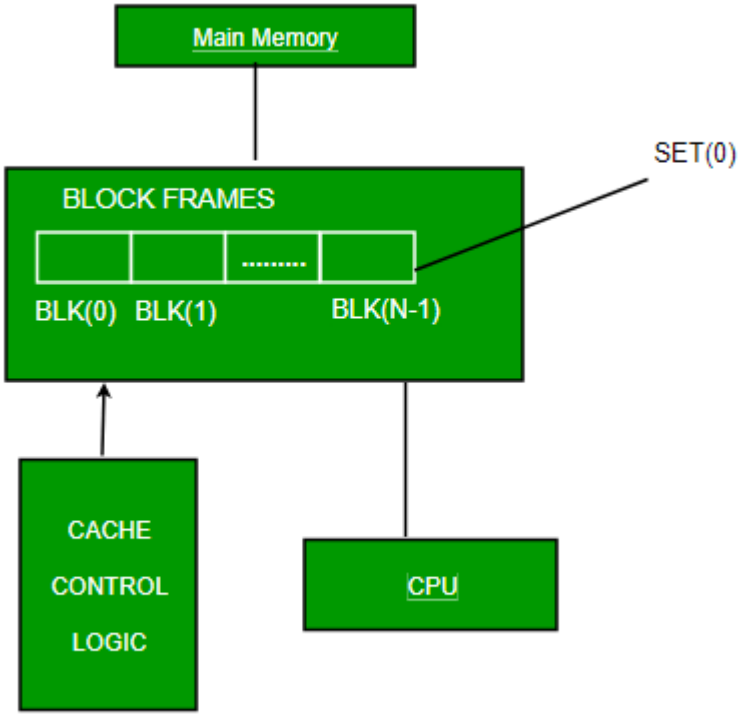
```
2. i = j modulo m

3. where

4. i=cache line number

5. j= main memory block number

   m=number of lines in the cache
```

   For purposes of cache access, each main memory address can be viewed as consisting of three fields. The least significant w bits identify a unique word or byte within a block of main memory. In most contemporary machines, the address is at the byte level. The remaining s bits specify one of the 2^s blocks of main memory. The cache logic interprets these s bits as a tag of s-r bits (most significant portion) and a line field of r bits. This latter field identifies one of the $m=2^r$ lines of the cache.



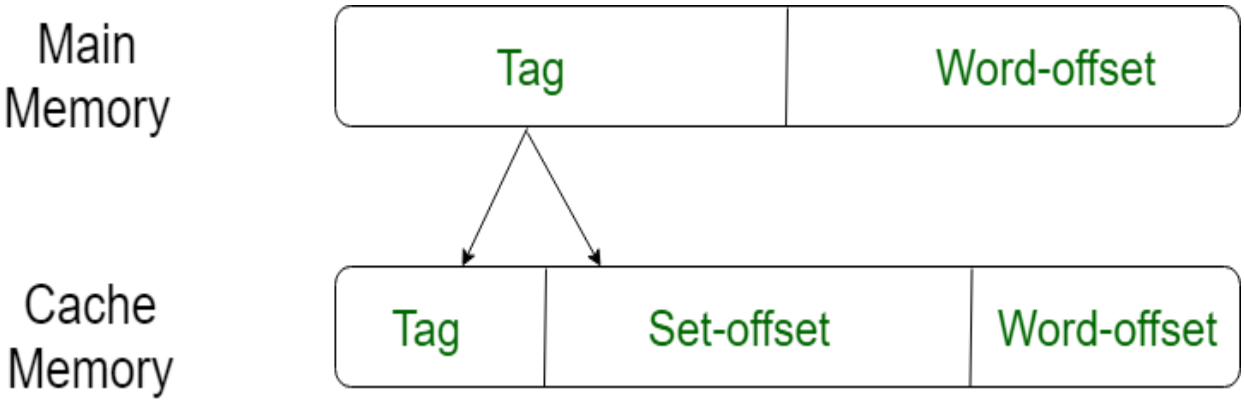


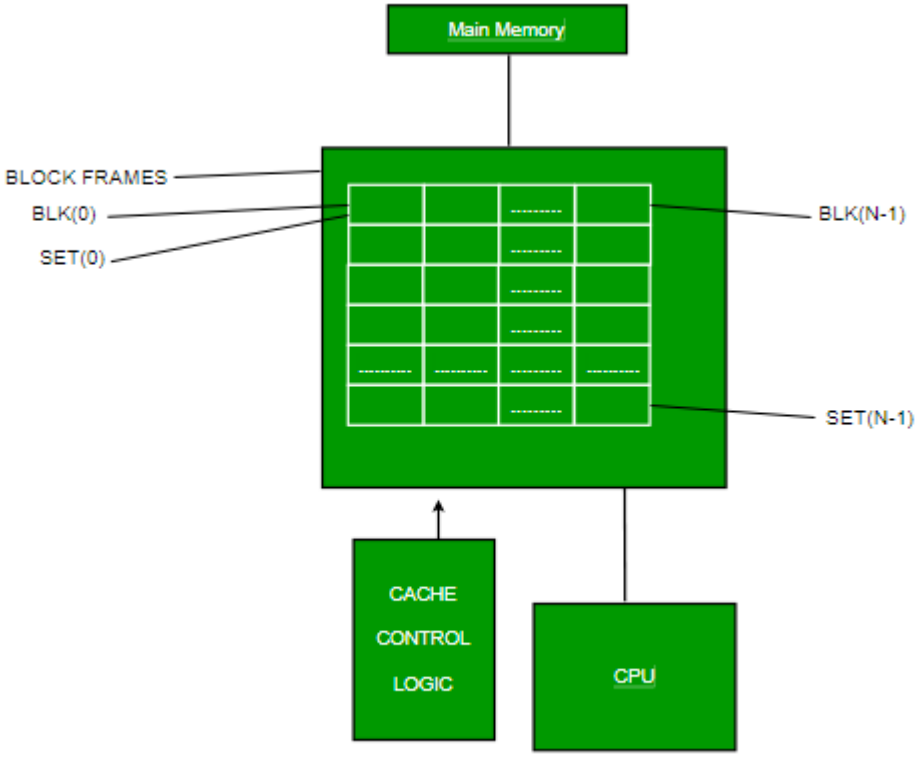
6. **Associative Mapping –**
   In this type of mapping, the associative memory is used to store content and addresses of the memory word. Any block can go into any line of the cache. This means that the word id bits are used to identify which word in the block is needed, but the tag becomes all of the remaining bits. This enables the placement of any word at any place in the cache memory. It is considered to be the fastest and the most flexible mapping form.



7. **Set-associative Mapping –**
   This form of mapping is an enhanced form of direct mapping where the drawbacks of direct mapping are removed. Set associative addresses the problem of possible thrashing in the direct mapping method. It does this by saying that instead of having exactly one line that a block can map to in the cache, we will group a few lines together creating a ***set***. Then a block in memory can map to any one of the lines of a specific set..Set-associative mapping allows that each word that is present in the cache can have two or more words in the main memory for the same index address. Set associative cache mapping combines the best of direct and associative cache mapping techniques.
   In this case, the cache consists of a number of sets, each of which consists of a number of lines. The relationships are

   ```
   m = v * k

   i= j mod v


   where

   i=cache set number

   j=main memory block number

   v=number of sets

   m=number of lines in the cache number of sets

   k=number of lines in each set
   ```





Application of Cache Memory –

1. Usually, the cache memory can store a reasonable number of blocks at any given time, but this number is small compared to the total number of blocks in the main memory.
2. The correspondence between the main memory blocks and those in the cache is specified by a mapping function.

Types of Cache –

- **Primary Cache –**
  A primary cache is always located on the processor chip. This cache is small and its access time is comparable to that of processor registers.
- **Secondary Cache –**
  Secondary cache is placed between the primary cache and the rest of

the memory. It is referred to as the level 2 (L2) cache. Often, the Level 2 cache is also housed on the processor chip.

Locality of reference –
Since size of cache memory is less as compared to main memory. So to check which part of main memory should be given priority and loaded in cache is decided based on locality of reference.
Types of Locality of reference

5. **Spatial Locality of reference**
   This says that there is a chance that element will be present in the close proximity to the reference point and next time if again searched then more close proximity to the point of reference.

6. **Temporal Locality of reference**
   In this Least recently used algorithm will be used. Whenever there is page fault occurs within a word will not only load word in main memory but complete page fault will be loaded because spatial locality of reference rule says that if you are referring any word next word will be referred in its register that's why we load complete page table so the complete block will be loaded.

Q3. What are the techniques for assessing the organization's environment and define the method of forecasting?

Ans) Human Resource Management performs organizational assessments for departments and schools. An assessment can help you identify key issues and opportunities that can improve your organization's overall effectiveness. We utilize a variety of tools to conduct our evaluations, including interviews and surveys.
This is a planned systematic review of an organization's processes, work environment, and organizational structure. The organizational assessment process guides the development of recommendations and action plans to support achievement of organizational objectives.
The scope and purpose of the assessment will help define how it is accomplished. The team will explore the organization's outcome by reviewing productivity and climate, strategic plans, goals and objectives, organizational charts, and operating procedures. Interviews with employees may also be conducted, as necessary.

Methods of Forecasting:
There are four main types of forecasting methods that financial analysts use to predict future revenues, expenses, and capital costs for a business. While there are a wide range of frequently used quantitative budget forecasting tools, in this article we focus on the top four methods:

(1) Straight-Line

(2) Moving Average
(3) Simple Linear Regression, And
(4) Multiple Linear Regression.

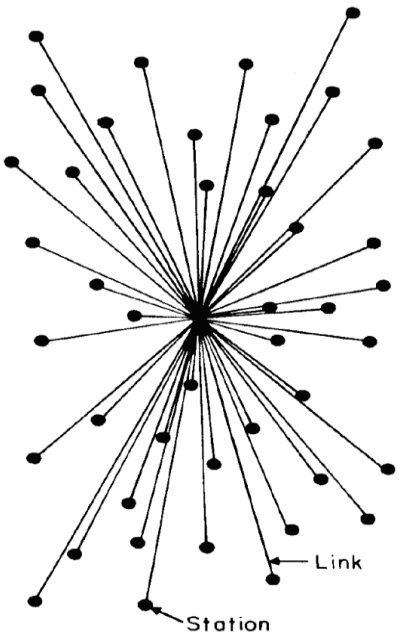
Q4. Differentiate centralize and decentralize arbitration.

Ans) Centralized systems may have helped build the internet, but they have important disadvantages. That's what decentralized try to address.

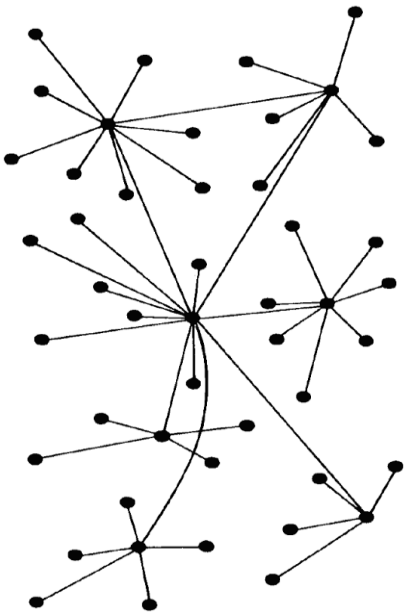
The Importance of Both Systems

The centralized vs decentralized systems debate is relevant to both individuals and organizations. It affects almost everyone who uses the web. It's at the core of the development and evolution of networks, financial systems, companies, apps, web services, and more.

The internet itself is the world's largest network. So large in fact that it brings together all these different systems into a vast digital ecosystem. But for most organizations and individuals, using all these systems is not feasible. They have to choose. And you may have to choose, too.



CENTRALIZED
(A)



DECENTRALIZED
(B)

Centralized Systems

In a centralized system, all users are connected to a central network owner or “server”. The central owner stores data, which other users can access, and also user information. This user information may include user profiles, user-generated content, and more. A centralized system is easy to set up and can be developed quickly.

But this system has an important limitation. If the server crashes, the system no longer works properly and users cannot access the data. Because a centralized system needs a central owner to connect all the other users and devices, the availability of the network depends on this owner. Add to that the obvious security concerns that arise when one owner stores (and can access) user data, and it’s easy to understand why centralized systems are no longer the first choice for many organizations.

Pros

- Simple deployment
- Can be developed quickly
- Affordable to maintain
- Practical when data needs to be controlled centrally

Cons

- Prone to failures
- Higher security and privacy risks for users
- Longer access times to data for users who are far from the server

Decentralized Systems

As its name implies, decentralized systems don’t have one central owner. Instead, they use multiple central owners, each of which usually stores a copy of the resources users can access.

A decentralized system can be just as vulnerable to crashes as a centralized one. However, it is by design more tolerant to faults. That’s because when one or more central owners or servers fail, the others can continue to provide data access to users.

Resources remain active if at least one of the central servers continue to operate. Usually, this means that system owners can repair faulty servers and address any other problems while the system itself continues to run as usual.

Server crashes in a decentralized system may affect the performance and limit access to some data. But in terms of overall system uptime, this system offers a big improvement over a centralized system.

Another advantage of this design is that the access time to the data is often faster. That's because owners can create nodes in different regions or areas where user activity is high.

However, decentralized systems are still prone to the same security and privacy risks to users as centralized systems. And while their fault tolerance is higher, this comes at a price. Maintaining a decentralized system is usually more expensive.

Pros

- Less likely to fail than a centralized system
- Better performance
- Allows for a more diverse and more flexible system

Cons

- Security and privacy risks to users
- Higher maintenance costs
- Inconsistent performance when not properly optimized

Q4. Elaborate the following 10 marks

a. Snoop

b. Snap

c. Locality of reference

d. Cache controller

Snoop:

A snoop filter is a directory-based structure and monitors all coherent traffic in order to keep track of the coherency states of cache blocks. It means that the snoop filter knows the caches that have a copy of a cache block. ... However, the exclusive snoop filter monitors the absence of cache blocks in caches.

Locality of reference:

Locality of reference, also known as the principle of locality, is the tendency of a processor to access the same set of memory locations repetitively over a short period of time. There are two basic types of reference locality – temporal and spatial locality.

Cache controller:

The cache controller designed here consists of four operations i.e. fetching address from the processor, read cache and main memory, write main memory and cache and providing the required data to the processor. All these operations are implemented using a Finite State Machine.

Snap:

The semantic network array processor (SNAP), a highly parallel architecture targeted to artificial intelligence applications, and in particular natural language understanding, is presented. The knowledge is represented in a form of the semantic network. The knowledge base is distributed among the elements of the SNAP array, and the processing is performed locally where the knowledge is stored. A set of powerful instructions specific to knowledge processing is implemented directly in hardware. SNAP is packaged into 256 custom-designed chips assembled on four printed circuit boards and can store a 16 K node semantic network. SNAP is a marker propagation architecture in which the movement of markers between cells is controlled by propagation rules. Various reasoning mechanisms are implemented with these marker propagation rules.