Course on E-Business
Professor Mamata Jenamani
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur
Module 05
Lecture Number 26
Hardware And Software Resources

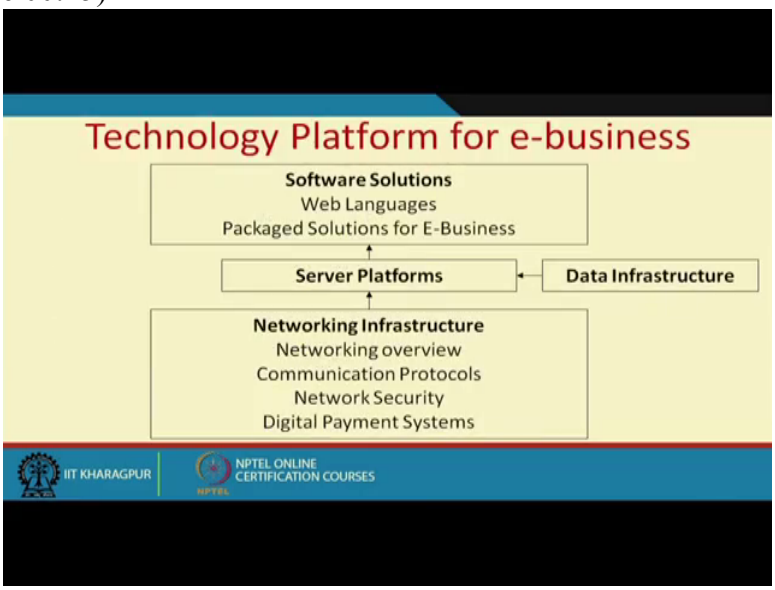
(Refer Slide Time 00:19)

So far we have been talking about the networking resources.

(Refer Slide Time 00:24)

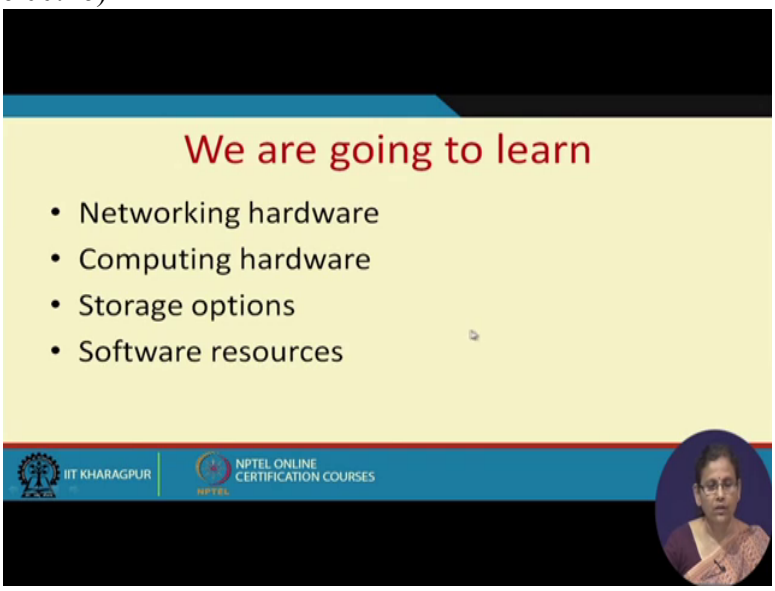
So if we look at

(Refer Slide Time 00:25)



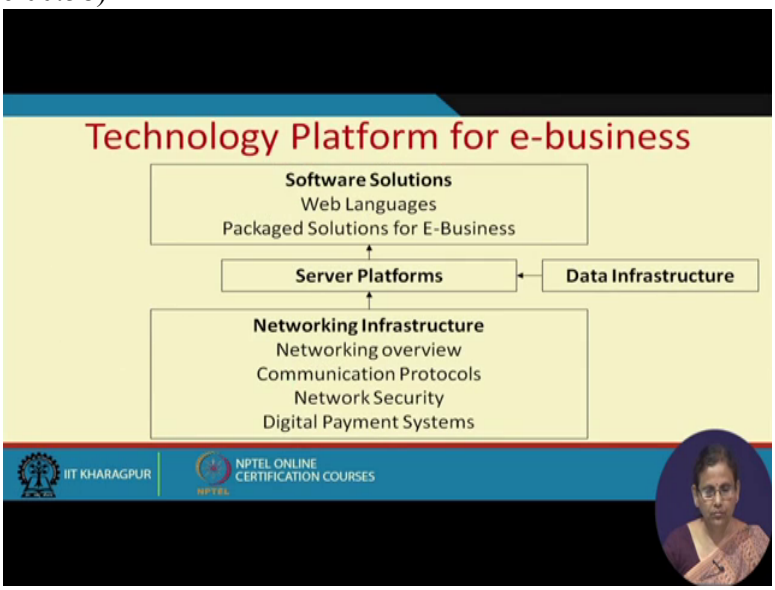
the technology platform for e –business we have, in the networking infrastructure, we have this networking protocols and along with that we have networking hardware. So when, now talking about this hardware, first thing that we are going to talk about

(Refer Slide Time 00:46)

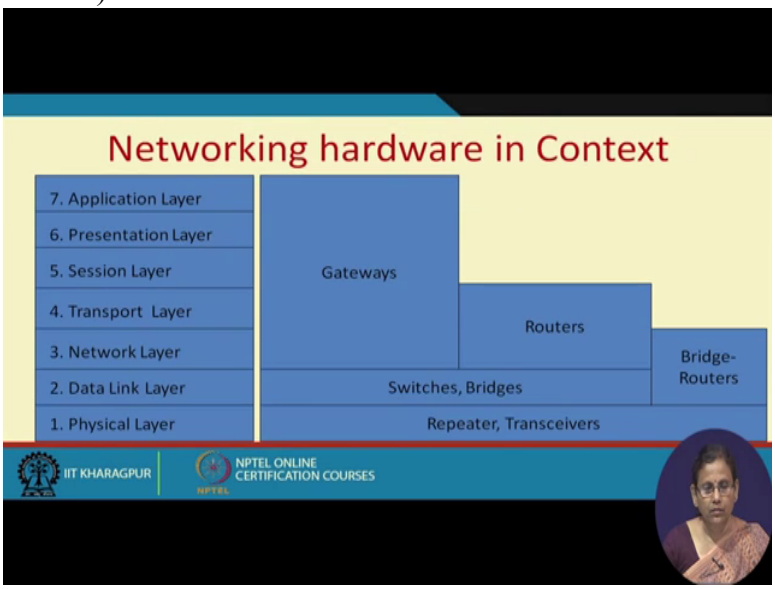


your networking hardware, then we would be talking about computer hardware, various storage options, then the software resources.

(Refer Slide Time 00:58)



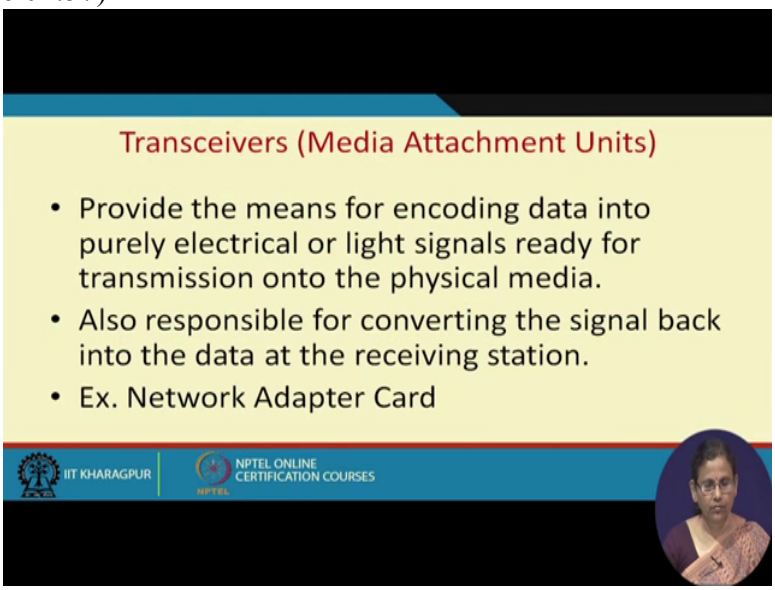
(Refer Slide Time 01:01)



Now if you remember our I S O/ O S I model which consists, which is a 7 layer model and last class we also mapped it to T C P/ I P. Now let us see what are various hardware devices in each layer of the protocol.

So first thing is that in the physical layer, you have repeaters and transceivers. Then in the data link layer you have switches and bridges. Then in the transport layer, you have...in the network layer you have, in your network and transport layer, you have your routers, then for upper layers you have gateways, then there are certain bridges to routers.

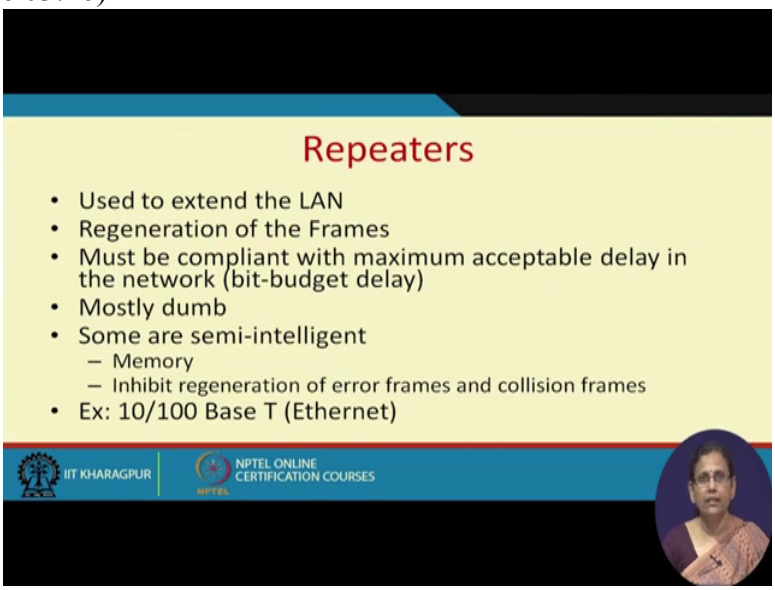
(Refer Slide Time 01:57)



Let us try to see what all of them, what is the functionality of each of them.

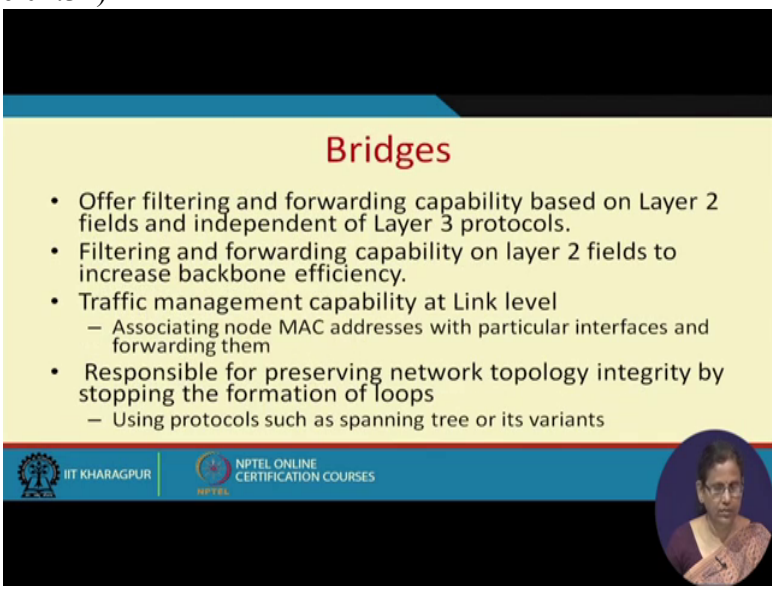
So first one is a transceiver. It provides the means for encoding the data into purely electrical and light signals and which is ready for transmission into the physical media. Now look you have different kinds of physical media. It can be coaxial cable, it can be optical fiber, optical fiber cable so depending on the type of physical media, the signal needs to be changed, signal is basically what, the data packets that you are sending ultimately has to be converted either to the electrical signal or to the light signal depending on the media or it can be converted, if it is a wireless protocol we are not right now not discussing then it has to be converted to appropriate form so that through the media it can go. So it is the transceivers who does it. Then they are also responsible for converting signals back into the data at the receiving stations. Your network adapter cards are actually examples of transceivers.

(Refer Slide Time 03:26)



Then the other element in the network is your repeater. It is used; these repeaters are used to extend the L A N. If as we have already discussed, during while passing through the media, there can be data loss. The packets may be lost or they may be, after all they are signals. So they may be lost or they may be weak. They may be distorted. So the repeaters are added within the network in, if the distance is more, to regenerate the frame. So they must be compliant with maximum acceptable delay in the network. They are mostly not very intelligent but some are semi-intelligent. They have memory and they can help preventing the errors in the frame and the collision of the frame. Ethernet is your one example of this.

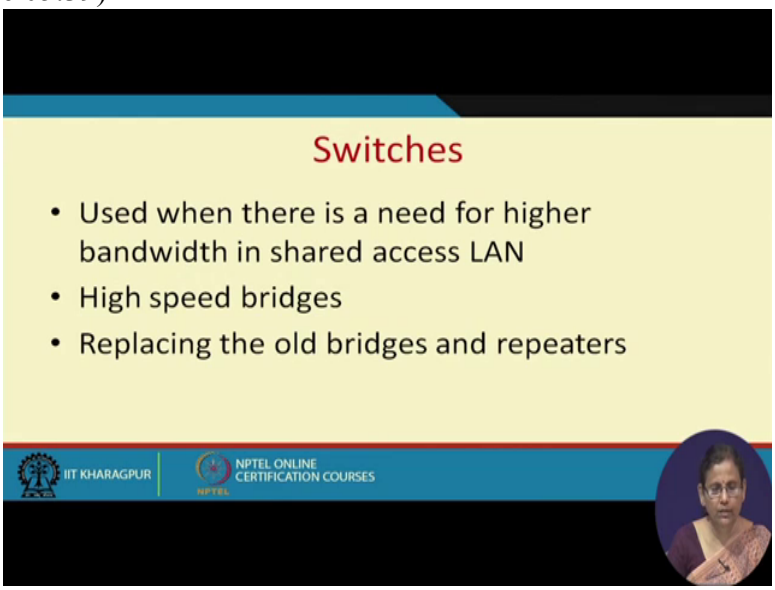
(Refer Slide Time 04:32)



Then you have bridges. These bridges offer filtering and forwarding capability based on layer 2 fields and independent of the layer 3 protocol. These filter, I mean they help in filtering and

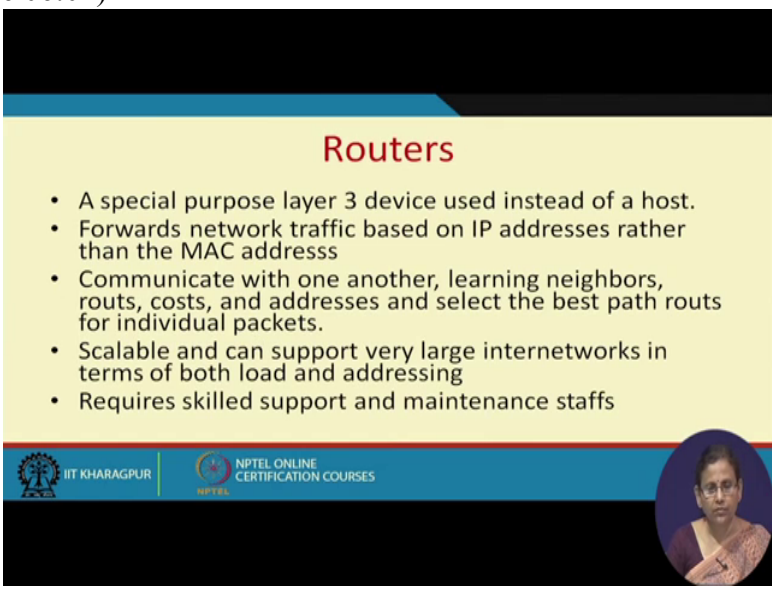
forwarding, they have this filtering and forwarding capability on layer 2 fields to increase the backbone efficiency. This traffic management capability, they have the traffic management capability at the link layer. They can be, they can associate the node to the corresponding MAC addresses with particular interfaces and forward them. They are responsible for preserving network topology and integrating by stopping the formation of loops. So such protocols, they use protocols such as spanning tree and its variant so that the frames which are sent over the internet are not come back to the same location.

(Refer Slide Time 05:39)



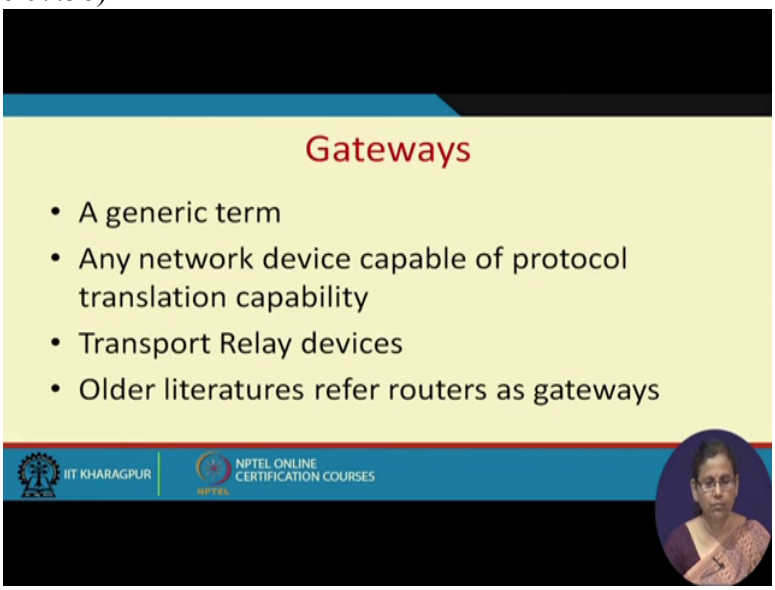
Then you have switches. They are used when there is a need for higher bandwidth in the shared access LAN. In fact these switches and bridges are, they are almost the same thing; they are high speed bridges. They are, I mean the replacing old bridges and repeaters.

(Refer Slide Time 06:02)



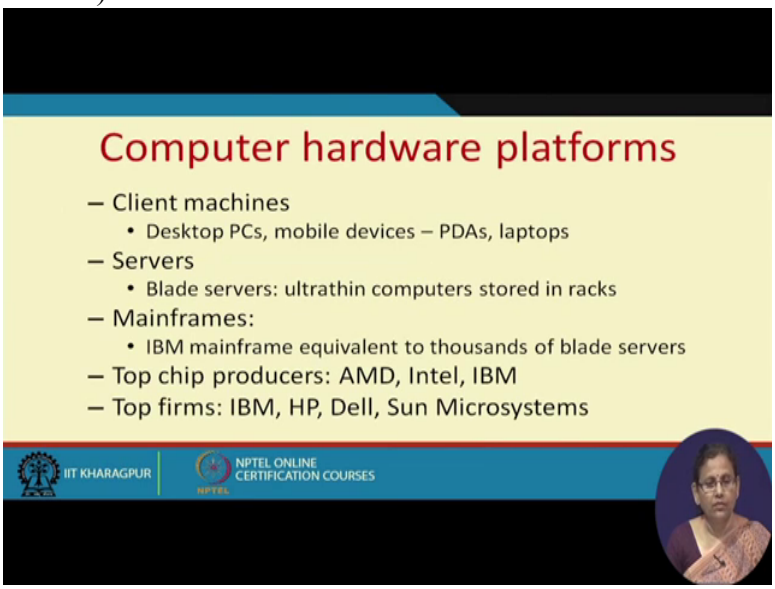
Then you have routers. These routers as I now told you when we look at the T C P/I P protocol stack, protocol stack the lower 2 layers are implemented in the routers. They are special purpose layer 3 devices used instead of a host. In fact while discussing about dynamic load balancing when we were discussing web based systems when the load becomes very high and you have multiple servers, that time also I mentioned that the routers are the devices. Sometimes they can map the I P address to Domain Name conversion, when it is used instead of the I P address of a particular host, I P address of a router is given which means each of the router has one I P address and they are intelligent enough to divert the traffic to a corresponding server in case of load balancing. Now this, these routers they forward network traffic based on I P address rather than based on MAC address. But the lower level it is the MAC address. Then they communicate with one another, learning the neighbors, routes, costs and addresses and select the best path routes for routing the individual packets. Now they are scalable and can support very large internetworks in both load and addressing; and they require skilled maintenance staffs to maintain them.

(Refer Slide Time 07:56)



Then gateway is a very generic term. Any network device that is capable of protocol translation when 2 different protocols, two different networks with, with different, with variant of protocols are connected, such devices are called gateways. They are just transport relay devices.

(Refer Slide Time 08:25)



Sometimes they are also mentioned as routers themselves.

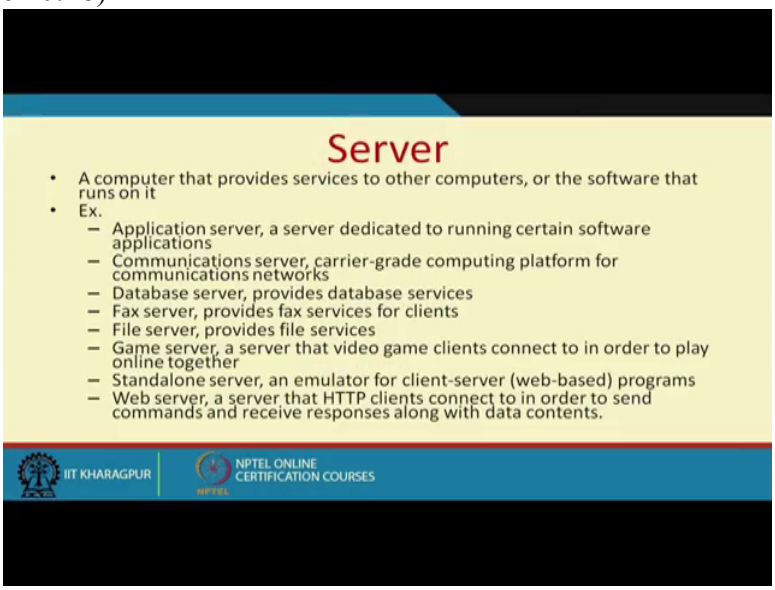
Now we are going to talk about the computer hardware platform. See, while

(Refer Slide Time 08:40)

talking about the resources the first thing that we have talked about is a server. Server is the basic element, by server we mean the hardware server, in fact in the earlier class I have told you when we talk about the server there is a hardware part of it, there is a software part. By hardware part we mean the, actually the physical server. Within that physical server you can have a web server or a H T T P server which is a software. You can also have a database server which also is a software. You can have, you also have an application server that is also a software. So the one that right now we are talking about is that hardware server or the hardware platform, the server that we are talking here is the hardware server.

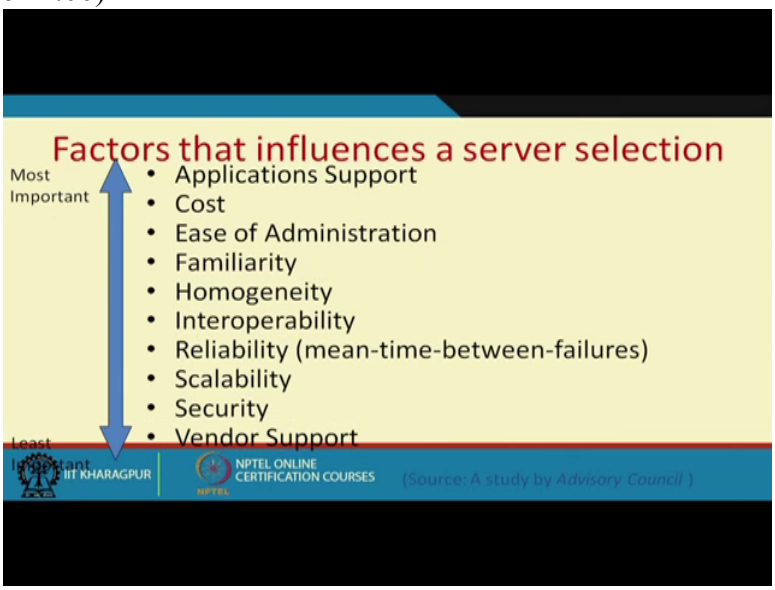
So different computing machines are part of your E- Business infrastructure. Your client machines which can be a P C, which can be a mobile device, which can be a laptop, similarly you can have servers which can be of again of many types, then you can have mainframes, then you can, all these servers contain the basic chips, the processors which are produced by the companies like Intel etc. And there are many firms which use these chip, the hardware chips to finally make the server, the companies like I B M, H P, Dell etc they make the server.

(Refer Slide Time 10:28)



Now a computer, what is a server, what is a computing server? The computer that provides services to other computer or the software that runs on it can be termed as the server. So server is the can be the hardware or it can be the software. So there can be many types of servers. There can be application servers, database server, your fax server and so on. They can be; all of them can be part of your business infrastructure.

(Refer Slide Time 11:00)

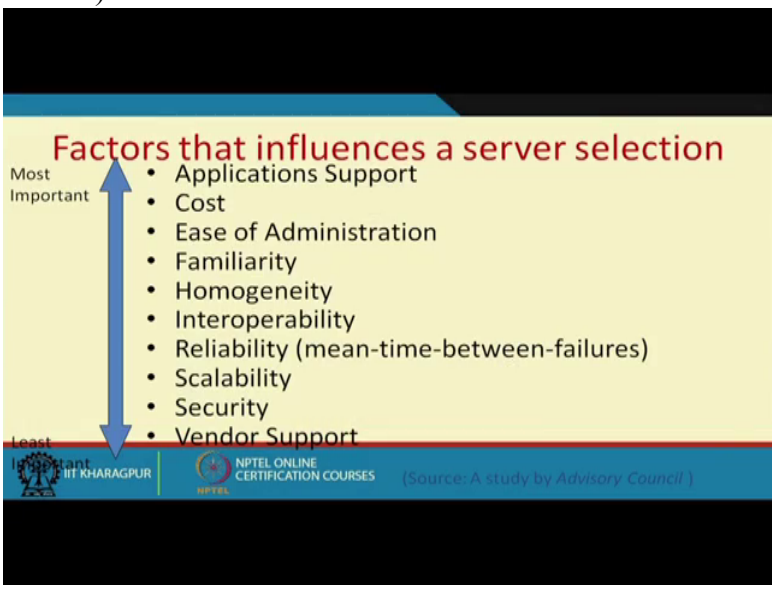


Then when you buy a server, many factors are actually responsible for this. We have arranged from most important to least important. First thing is that it should have, it should be supporting the application that your company requires. Second important

(Refer Slide Time 11:24)

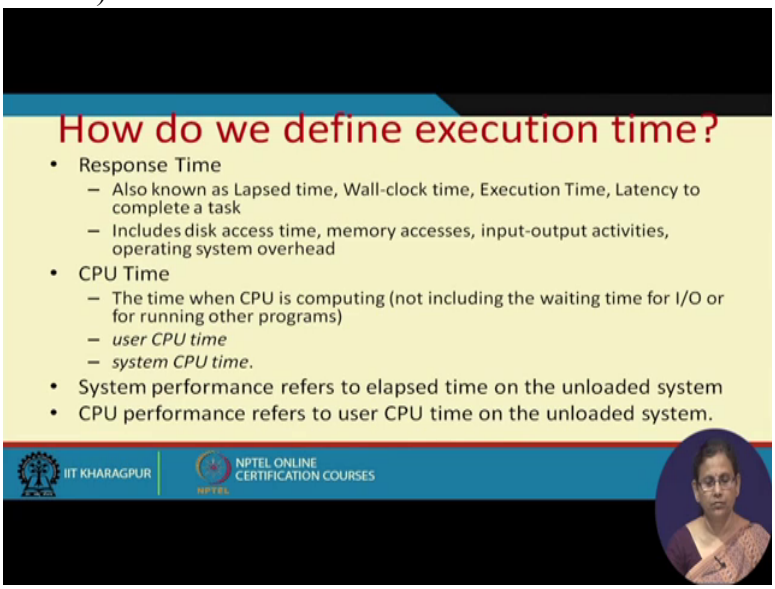
element is the cost. Third is the ease of administration.

(Refer Slide Time 11:30)



Fourth one is familiarity with that, with similar range of systems; homogeneity by homogeneity we mean if you are buying a new server, it should be actually be compatible with your existing systems, then interoperability, if there are 2, there are other systems in your organization or in your business partner's organizations then this particular server that you are buying should be providing you the support for interoperability. It should be reliable; it should be scalable with proper security et cetera. There are certain servers which are more prone to virus attack et cetera, certain servers are not. So then the another important fact is actually vendor support. Though it is of not very high priority, but this is of priority as well. If you have a good vendor support then your maintenance problem etc, etc can be taken care of.

(Refer Slide Time 12:47)



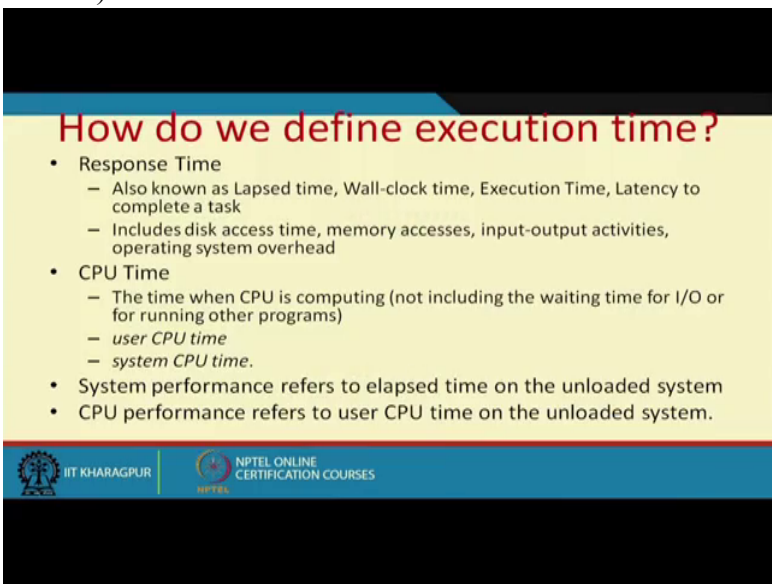
Now when we buy a server, in fact in the very first lecture in this series of Infrastructure in E-Business we understood that

(Refer Slide Time 13:04)

performance of a server can be measured through response time. Customers over the internet cannot really wait if your server is slow or your website is slow. And that time we discussed there are many factors responsible if a website is slow. Your server is responsible, your I S P can be responsible, your network can be responsible. Even your client machine can be responsible. But it is the server which is the most important, at least it is, we have the control over it and we should be buying a server with good response time.

Now what exactly

(Refer Slide Time 13:47)



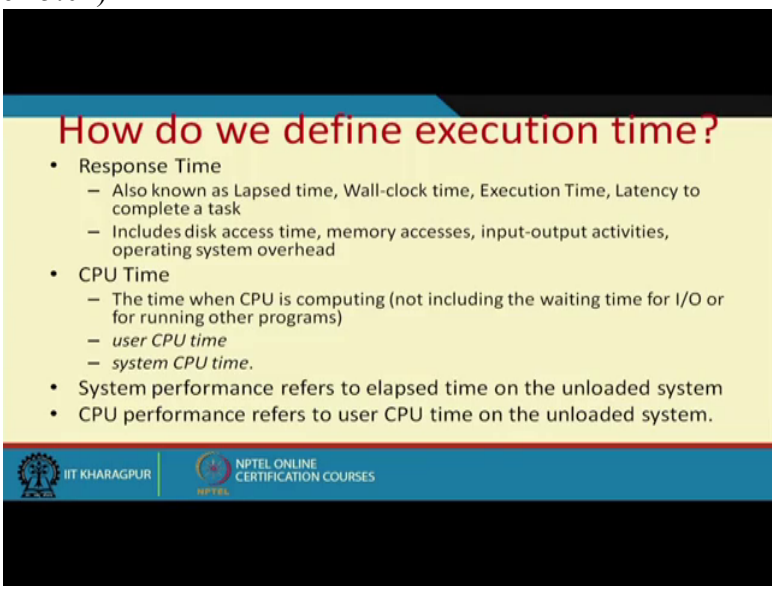
this response time is? This response time also known as the lapsed time, this lapsed time or wall-clock time or execution time or latency to complete a task, this includes the disk access time, memory access time, input output activities, operating system overhead and so on. So which means whenever your request is,

(Refer Slide Time 14:13)

just consider the example of a H T T P request. When H T T P request is sent to a web server, then the web server in turn will be contact, if it is a dynamic page, it might be, the web server will be contacting the application server. And the application server in turn will be contacting the database server.

Besides these server activities there will be operating system's involvement as well, operating system overhead. There will be frequent memory accesses. So all these things together define your execution time. Then within the computing devices, you have a C P U, the central processing unit

(Refer Slide Time 15:02)



and this is the main part for making the computations happen. So the C P U time is a major part of your response time. So your C P U should be powerful enough so that,

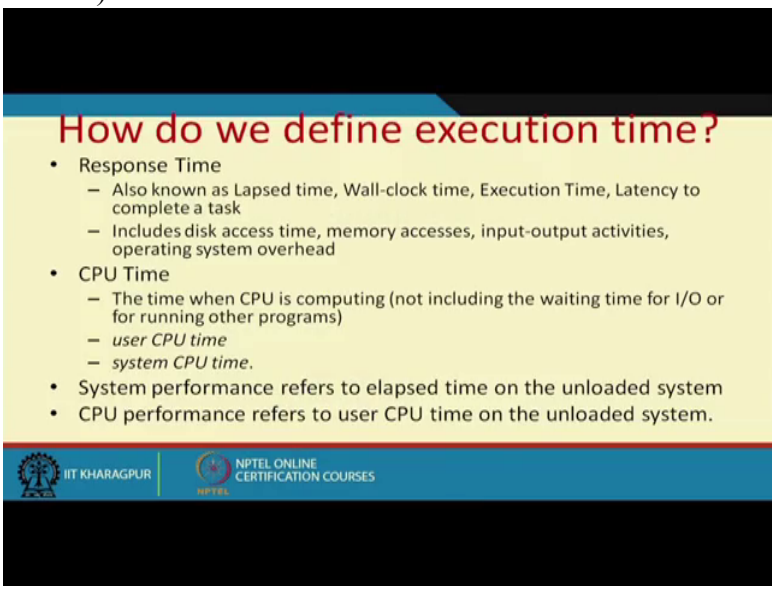
(Refer Slide Time 15:24)

and powerful and your operating system should be capable that C P U does not wait for I/ Os for running other programs. This C P U time can be broadly classified into either user C P U time for running a single program or the C P U, the system C P U time which includes not

only the program running time, it also includes the operating system overhead over running the program.

Now when you buy a system, buy specifically a server, your system's performance is referred to as the elapsed time

(Refer Slide Time 16:02)

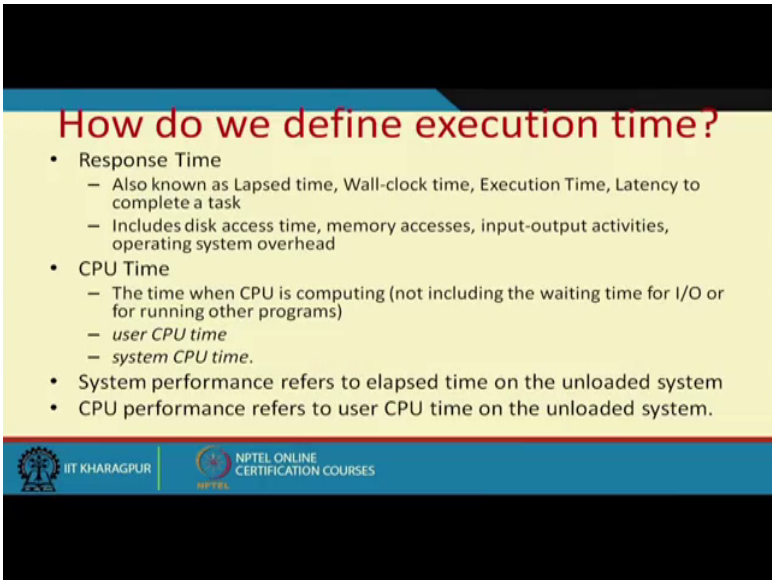


on the unloaded system. By unloaded system we mean no other

(Refer Slide Time 16:09)

program runs at that particular time. Only a specific

(Refer Slide Time 16:16)

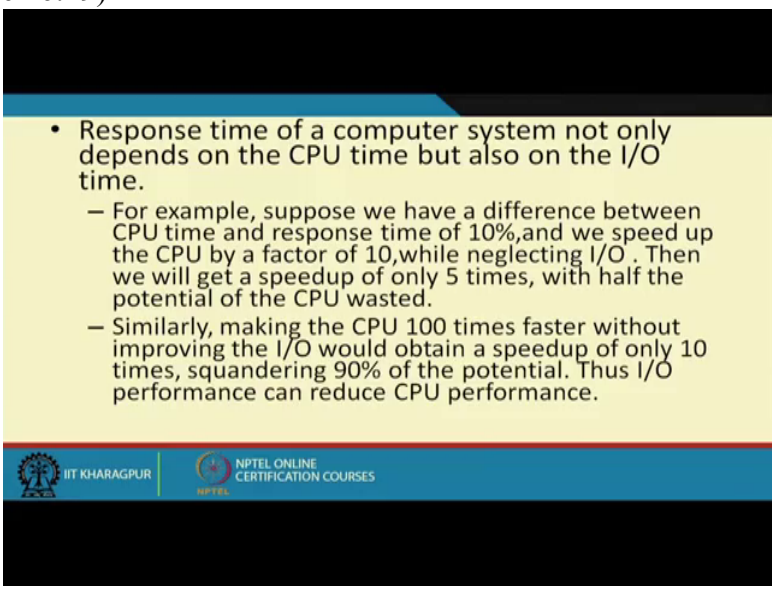


program for which the performance is to be measured runs. Then

(Refer Slide Time 16:22)

the C P U performance refers to the user C P U time on the unloaded system. Now

(Refer Slide Time 16:29)

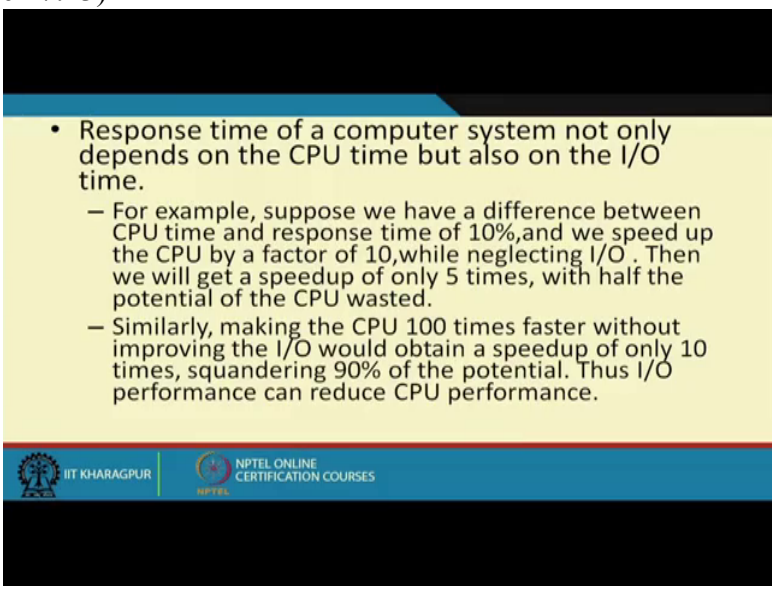


this response time of a system not only depends on the C P U time, but it also depends on the I/O time. By input output time we mean your system is connected to many input and output devices. So while making a choice of the server, only C P U time is not important. If a request, if a program is running and it in turn requests accessing the I/ O devices, it has to wait and if the I/ O devices are slow then that waiting time will be

(Refer Slide Time 17:11)

more. In fact it is seen that if you have, I mean suppose we have a difference between C P U time and response time of

(Refer Slide Time 17:23)



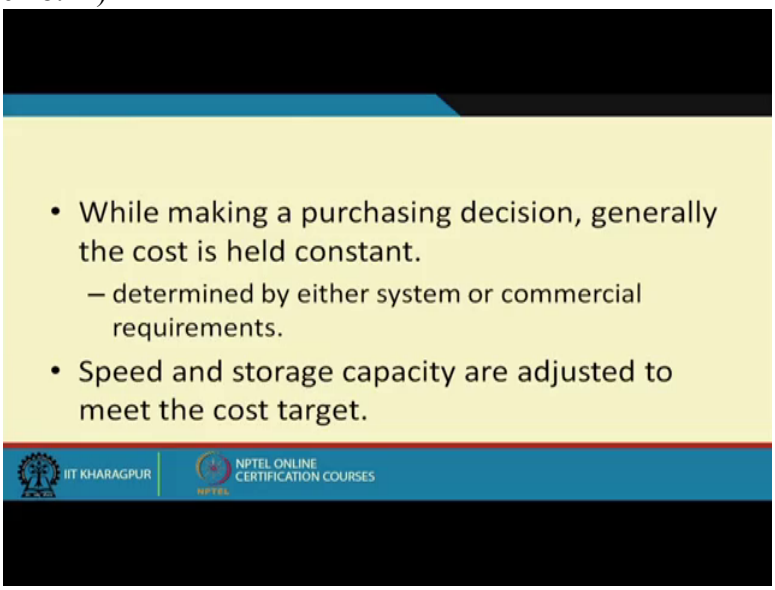
10% and we speed up C P U time by factor of 10 while neglecting the I/O we will get the speed of only 5 times more. It will not increase by 10 times, by 10% with half of the potential of the C P U wasted. Similarly making a C P U 100 times faster without improving the I/O would obtain the speed of only 10 times and wasting 90% of its potential. Thus I/ O performance can reduce the C P U's performance. So while

(Refer Slide Time 17:57)

deciding a server, it is not only the C P U's characteristics but its I/ O characteristics also need to be understood.

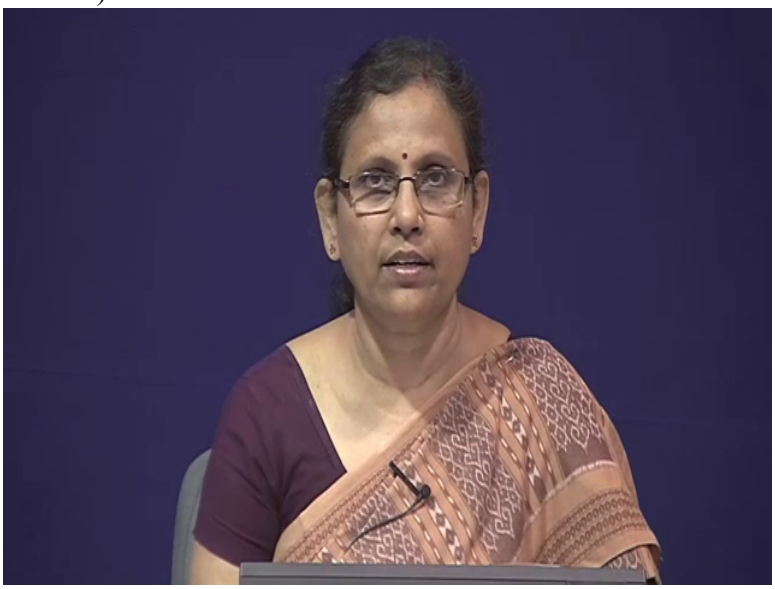
While making a purchasing decision, generally the cost is held constant because cost is

(Refer Slide Time 18:14)



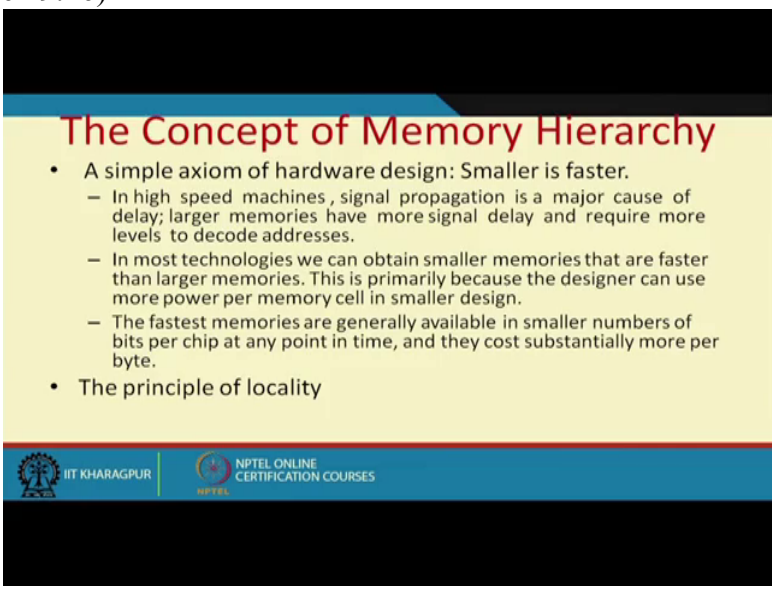
a major constraint as we have discussed that what are various parameters to choose. The second important parameter is actually the cost. So keeping the cost constant, the system's parameters are adjusted. You decide about various elements to be added to your system keeping this cost constant. The speed and storage capacities are adjusted to meet the cost target.

(Refer Slide Time 18:47)



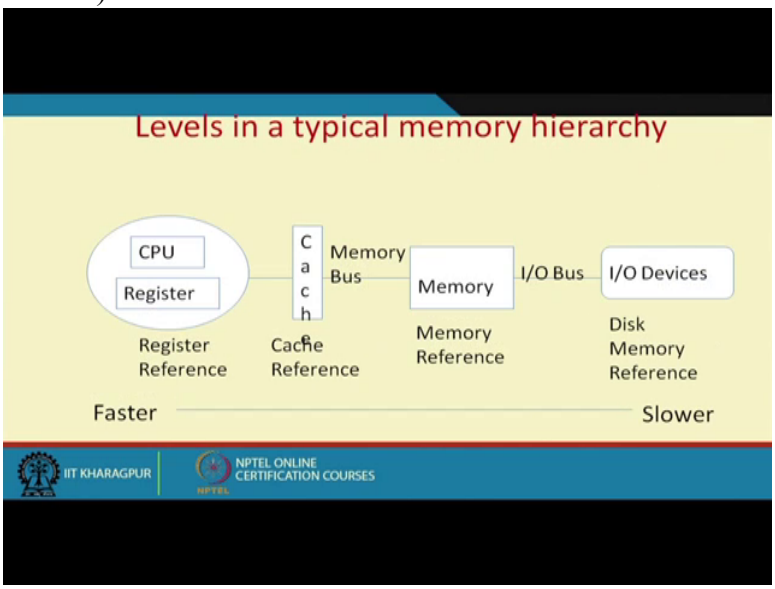
Now as I have told you it is important that you have a C P U, you have a high speed C P U but your connecting, supporting devices should also be fast. Specifically your memory plays a very important role. This memory is addressed is, actually realized in the,

(Refer Slide Time 19:16)



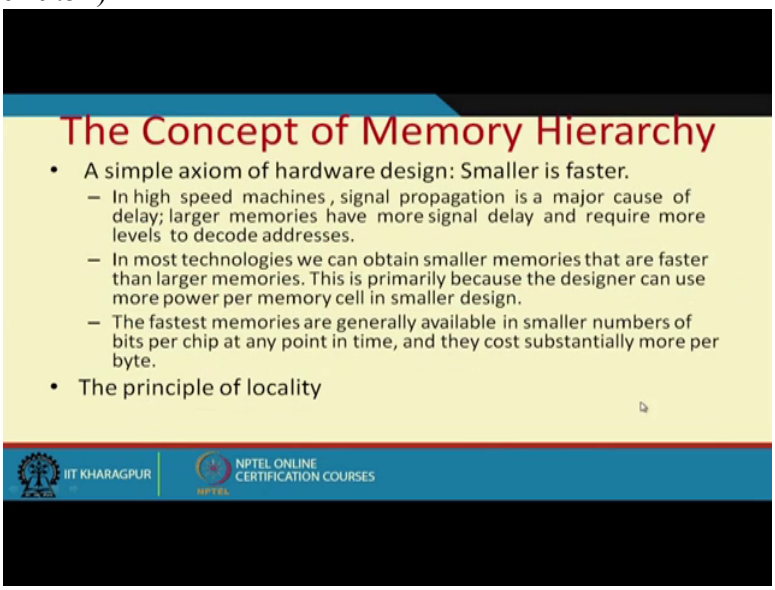
computer's memory in terms of memory hierarchy. In this hierarchy

(Refer Slide Time 19:22)



first you have, nearest to the C P U you have these registers. Then you have the cache, then you have the main memory. Then you have the disk memory. The fastest memory is near the C P U and the slowest memory is farthest from the C P U. So

(Refer Slide Time 19:54)



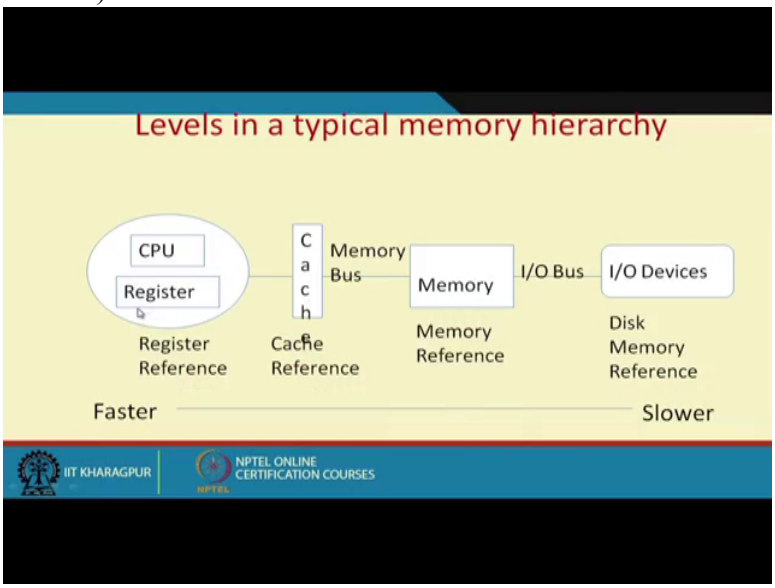
in high speed machines the signal propagation is a major cause of delay. Now larger memories have more signal delays and require more level, more levels to decode the addresses. Therefore to keep it simple, near the CPU, the fastest memory and the smallest memory is kept. In most technologies we can obtain smaller memories that are faster than the larger memories. This is primarily because designer can use the, use more power per memory cell in smaller design. Now fastest memories are generally available in small number of bits per chip at any point of time and they cost substantially more per byte. And they are kept nearer, the fastest memory elements are kept nearer to the C P U because,

(Refer Slide Time 20:56)

because due to this principle of locality, the same program code or the same blocks of data are going to be accessed, I mean the nearby blocks of the data or nearby parts of a instruction can be accessed immediately.

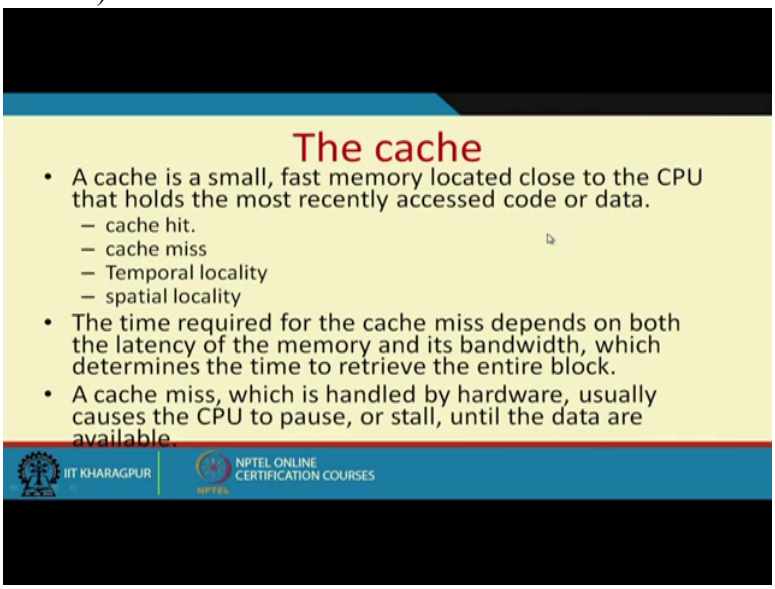
So the first element, let us have a look at this memory hierarchy once again.

(Refer Slide Time 21:30)



First is your CPU. Then is your cache memory. And C P U has its register. Then you have cache memory. Then you have main memory. Then you have your disc memory.

(Refer Slide Time 21:42)

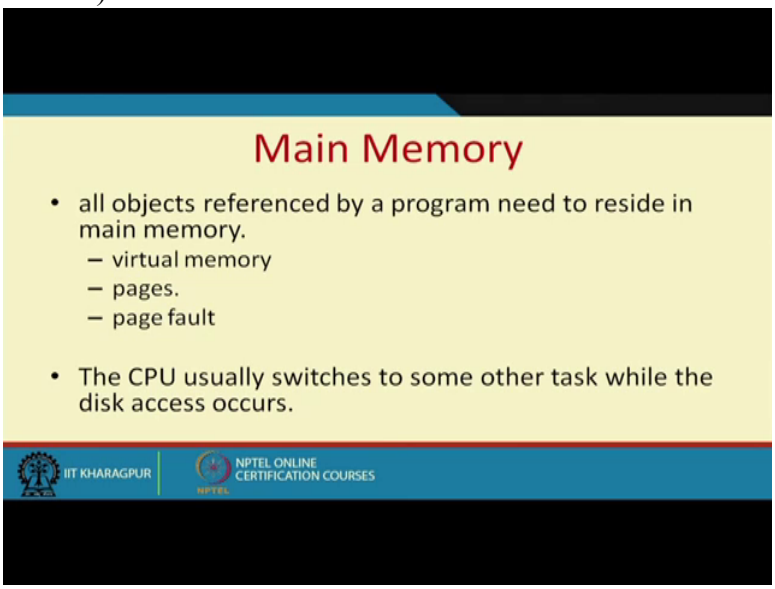


The cache memory is the smallest and fastest located closed to the C P U that holds the most recent address codes or the data. When the C P U, when some data is not there in the register when C P U searches, it first searches in the cache. This, if it is, the data is available, the data or instruction is available in the cache, it is called the cache hit. Otherwise it is a cache miss

because of this temporal locality. We were discussing about this principle of locality. This principle of locality can be temporal locality which means the data which is accessed now, in the near future, it is going to be accessed again. And this spatial locality says if I am getting, accessing a small chunk of data, after, and the data which is immediate to this, immediate to this, I am next going to access it. So the time required for cache miss depends on both the latency of memory and its bandwidth which determines the time to retrieve the entire block. This cache miss which is handled by the hardware is usually caused by, cause the C P U to pause or stall until the data is available. So therefore most used data has to be kept in the cache.

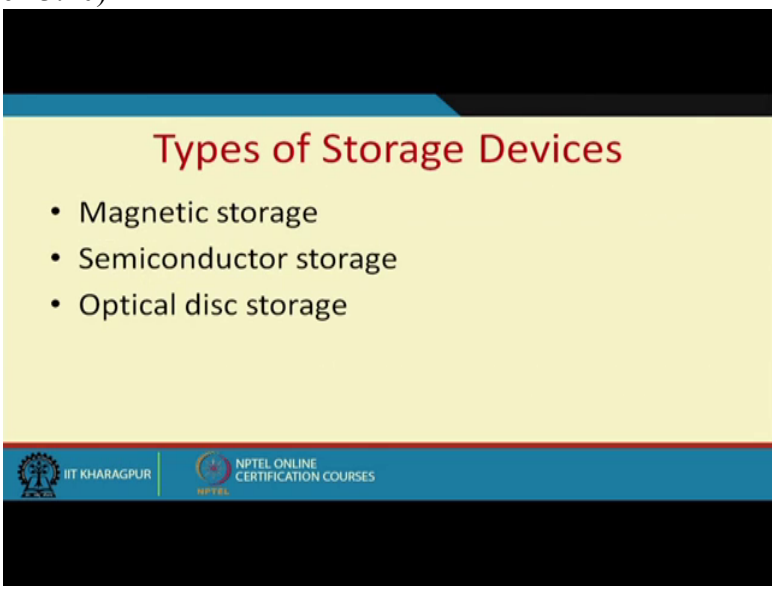
Now if the cache, the data is not available in

(Refer Slide Time 23:10)



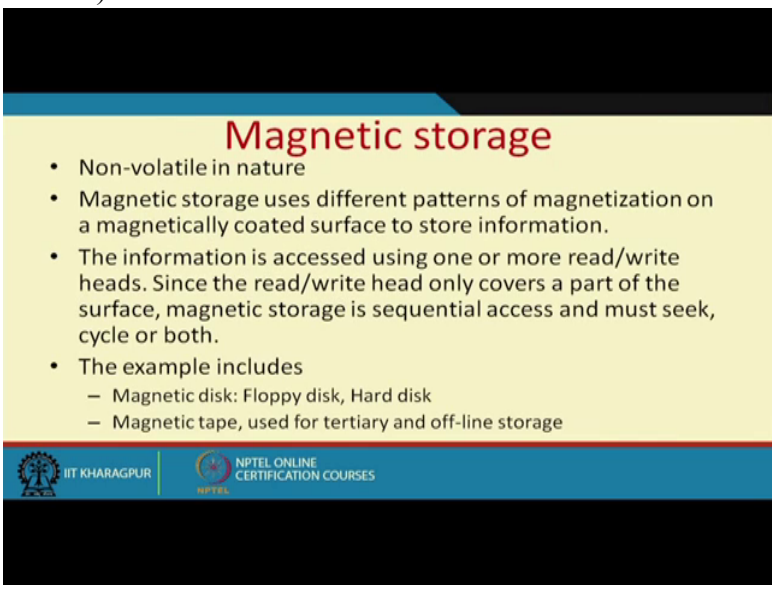
the cache memory, they will be going to the main memory. And if they are not in the main memory, if the main memory is, is not adequate then may be a part of the secondary memory can be used as a virtual memory. So if it is not available in your RAM or in your main memory then page fault occurs. And if the page fault occurs, then disk access has to be happen.

(Refer Slide Time 23:40)



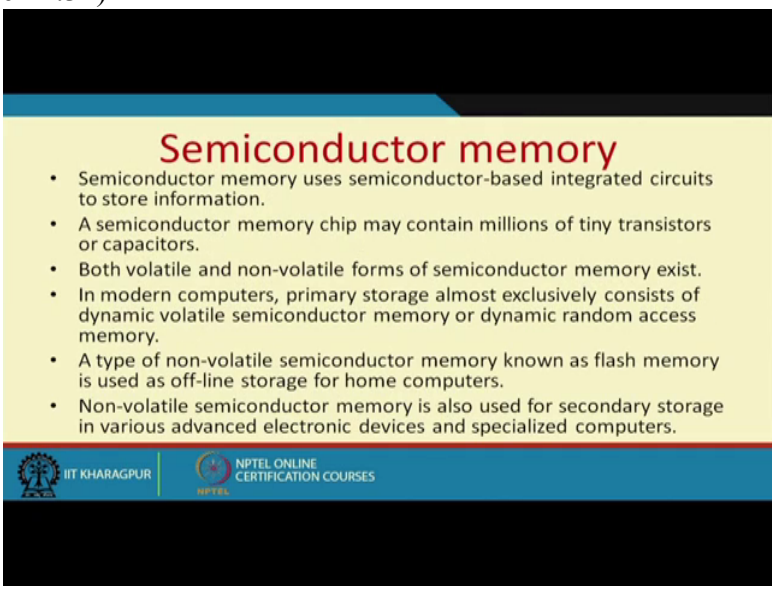
Now these secondary storages can be in, in various forms. First one is your magnetic storage. Second one is your semiconductor storage. It can be optical disc storage as well.

(Refer Slide Time 23:57)



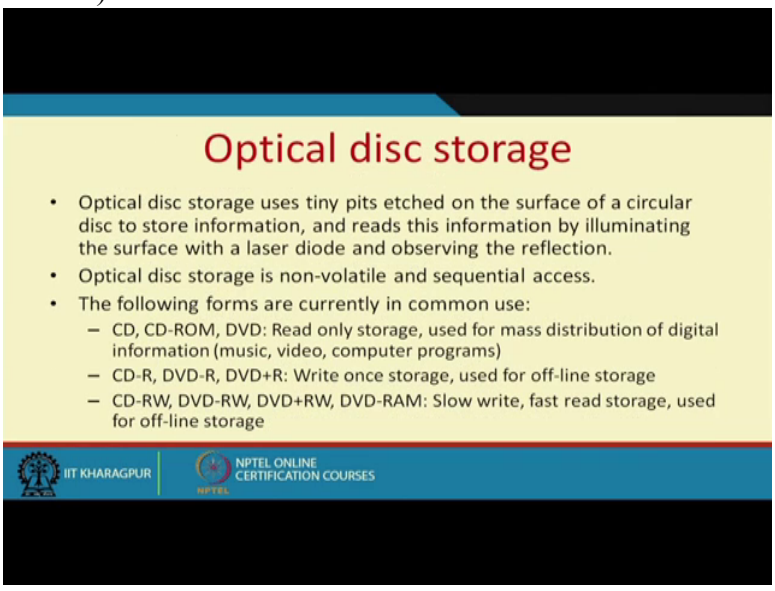
This magnetic storage is, your main memory while the main memory is volatile, all these secondary memories are non-volatile. This magnet also, so also this magnetic storage which can be magnetic disc, magnetic tape etc all of them are non-volatile in nature. And the data can be stored in them and can be accessed in future instances.

(Refer Slide Time 24:32)



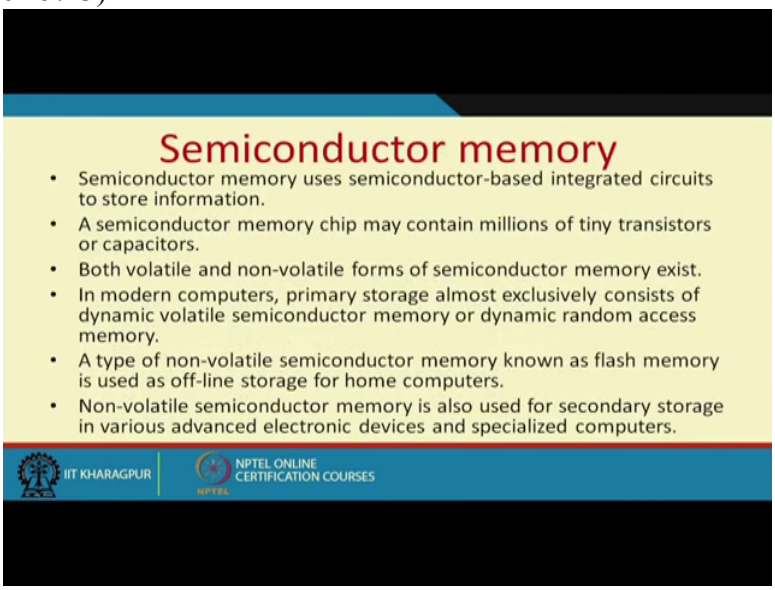
This semiconductor memory is also a permanent memory which contains millions of tiny transistors or capacitors. They come in both volatile and non-volatile form. In fact the volatile part of it is the, kept in the, as the main memory and non-volatile part, today you see your pen drive etc. they are actually semiconductor memory.

(Refer Slide Time 25:06)



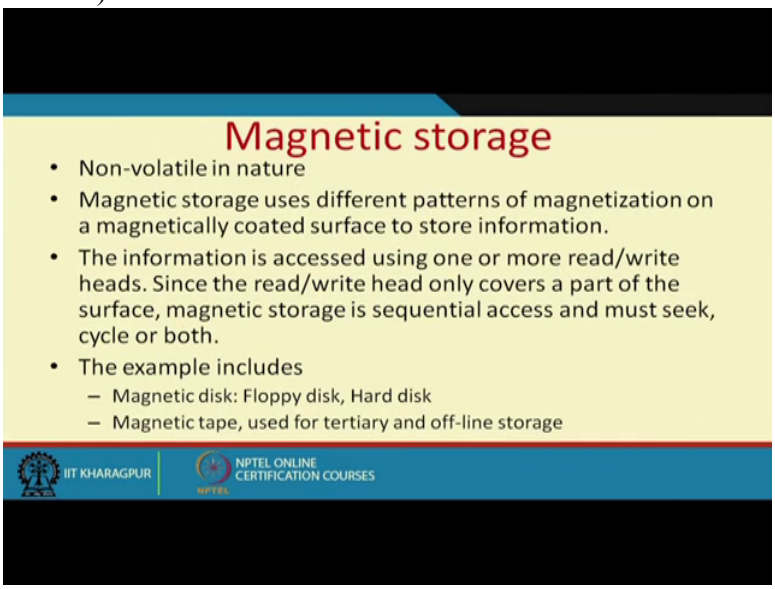
Then next is your optical disc storage which is your C D ROM etc. Ok while talking about these memories they actually differ in the way the data is stored in them. In case of magnetic

(Refer Slide Time 25:23)



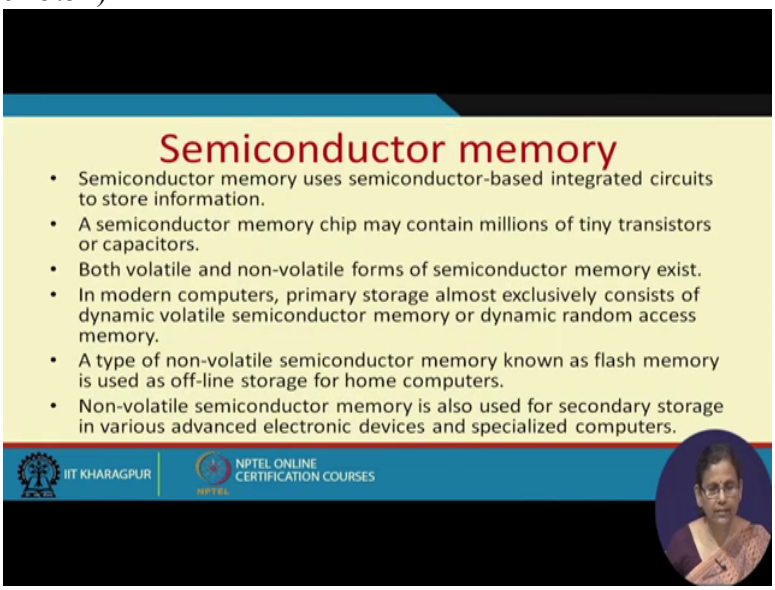
memories,

(Refer Slide Time 25:24)



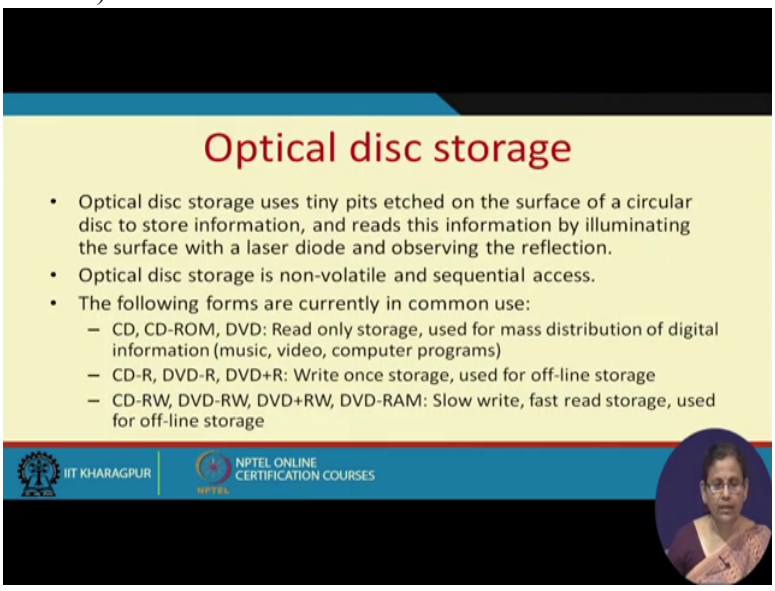
magnetic memories, the data is stored by the magnetic, I mean the different magnetization patterns, different magnetization patterns on a coated surface, on a surface which is coated by some magnetic material; where as

(Refer Slide Time 25:51)



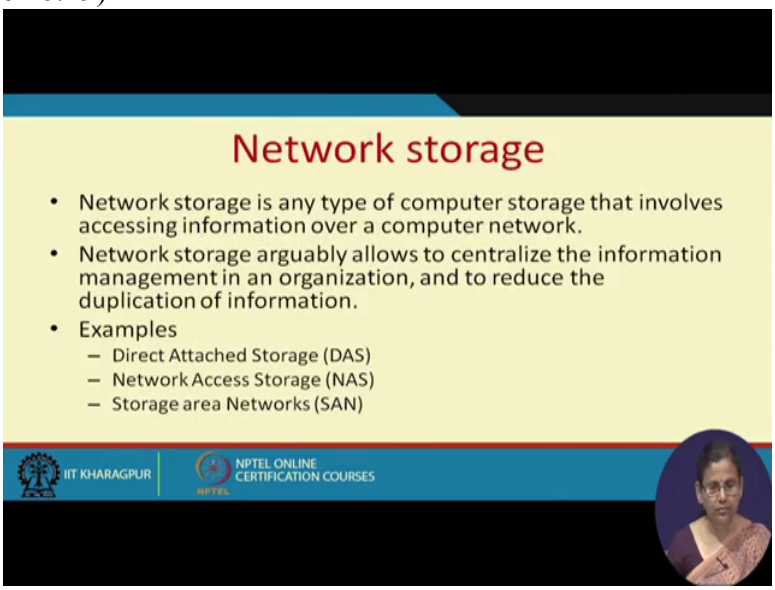
in case of semiconductor it is the transistors which are actually responsible for storing the data.

(Refer Slide Time 26:03)



In case of optical disc which is again non-volatile in nature, it is, the information is stored in a different manner. It is by illuminating the surfaces with laser diode and observing the reflection and while storing the data, it actually, I mean you must be knowing that you have to burn the C D, by burn the C D we mean we create different patterns which can be afterwards read using, by illuminating the surface.

(Refer Slide Time 26:49)

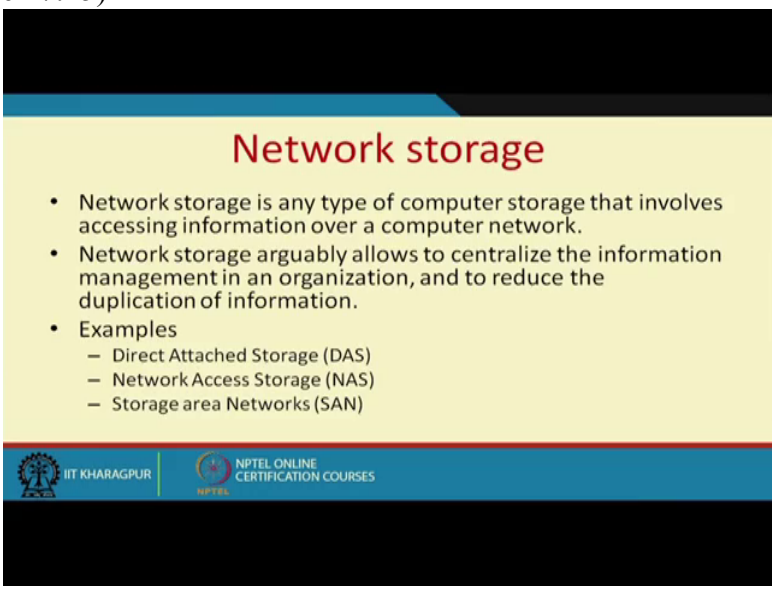


Then another category of storage devices which are actually responsible for keeping your organizational

(Refer Slide Time 26:57)

data is your network storage devices. They are not part of your, they may or may not be part of your computing, they are in fact not part of your computing system. They are separate memories

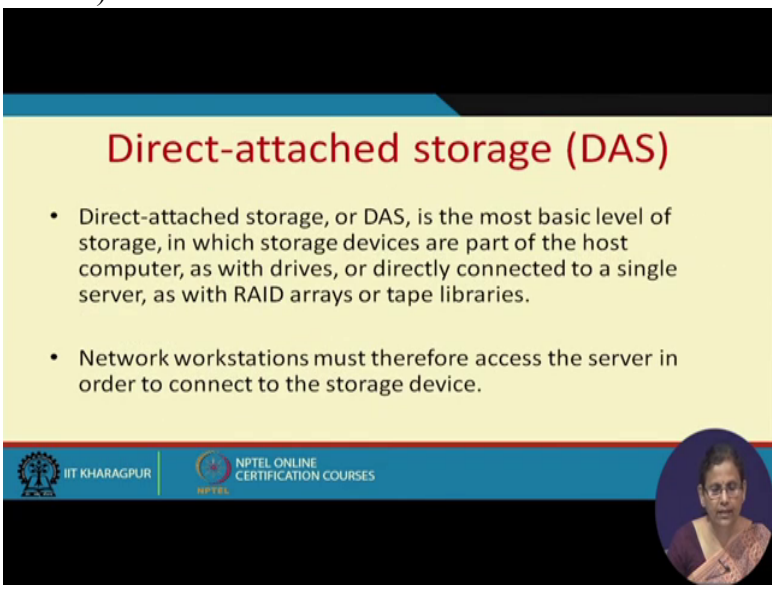
(Refer Slide Time 27:15)



to which your computer can access. This network storage is any type of computer storage that involves accessing the information over communication network. These network storages allow to have a centralized storage of information in the organization to reduce the duplication of, duplication of it.

There are 3 types, direct attached storage, network access storage and storage area network. In case of

(Refer Slide Time 27:51)



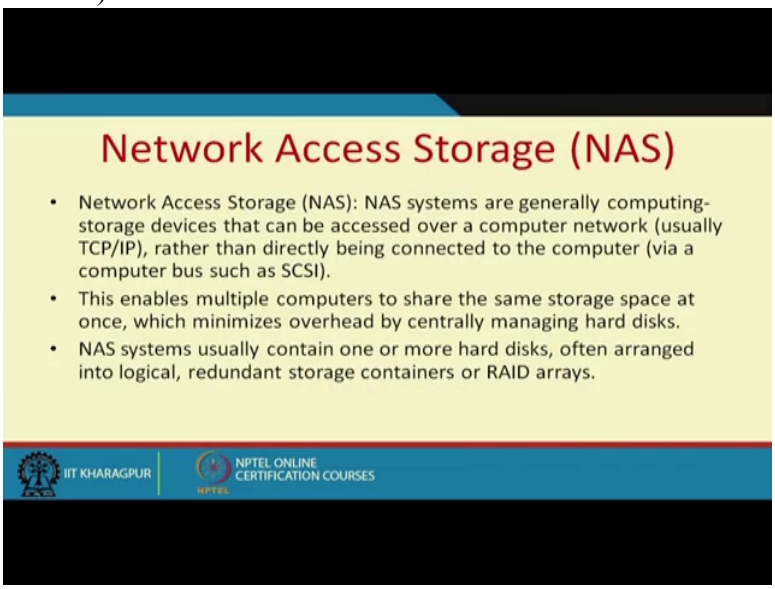
direct attached storage which is a basic form of storage, the storage devices are part of the host computer as drives and directly connected to a single server with these RAID arrays and all, about all these details we are not anyway going to discuss but they are additional

secondary memory, additional drive attached to a server. This network workstation must therefore access the server in order to connect to this storage device. Which means there will be a server, with that server this device will be attached. Other computers in the network can access the server first and through the server they can access this direct access storage device.

(Refer Slide Time 28:50)

This network storage, these network

(Refer Slide Time 28:53)

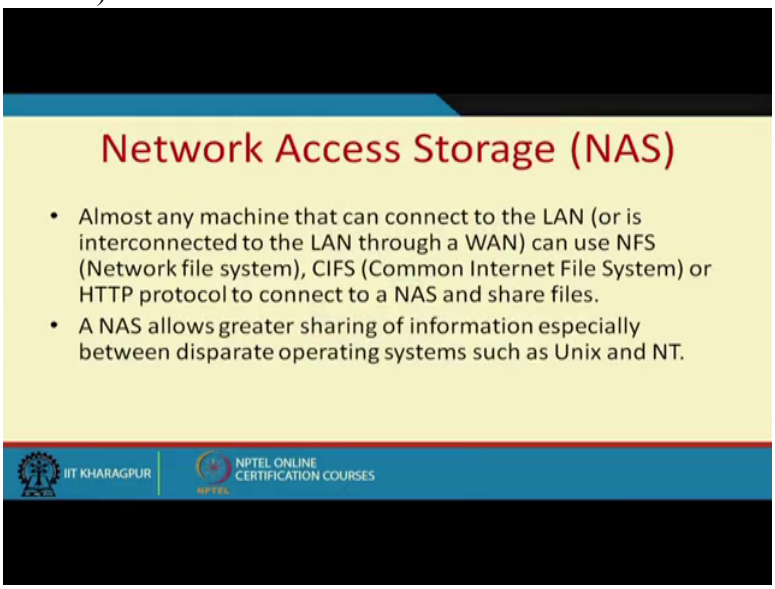


access storage devices in contrast to your direct access storage, they have their own I P addresses and other computing devices in your network, if they would like to access this storage device, they need not go to a specialized server which is in

(Refer Slide Time 29:14)

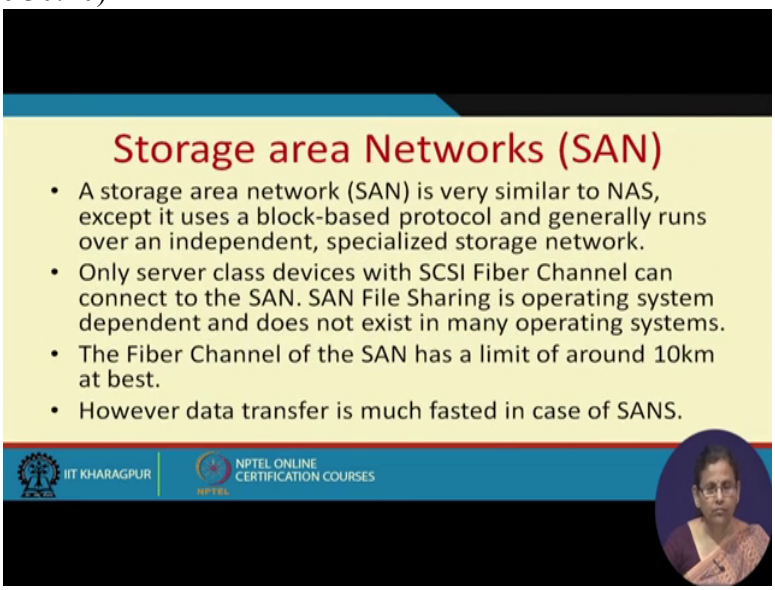
case of DAS, your direct access storage. So this enables multiple computers to share the same storage space at once which minimizes the overhead by centrally managing the hard disks. So NAS systems usually contain one or more hard disks and they are often managed in logical redundant storage containers or RAID arrays. Then almost any machine that

(Refer Slide Time 29:49)



can connect to LAN can use these network file system or this common internet file system or H T T P protocol to connect to these NAS devices and share files. NAS allows greater sharing of information specially between different kinds of, it has the capability of connecting to different kinds of operating systems as well.

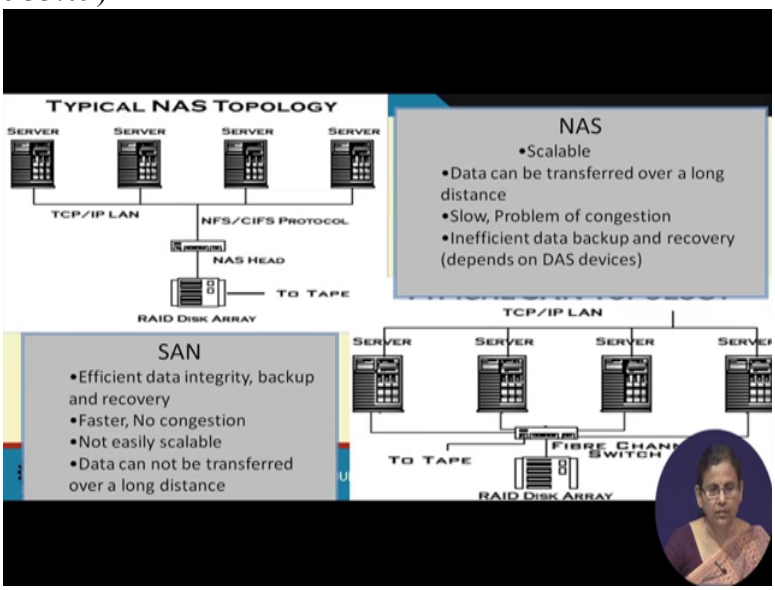
(Refer Slide Time 30:20)



Then the next category of network storage device is storage area network or SAN. This is very similar to the NAS, however the difference between the NAS and SAN is, NAS, you can, your computing devices can access NAS over the same local area network to which other computing devices are attached. So it has one I P address and everybody, all other computing devices can access NAS. However in case of SAN, that is storage area network, you have to have specialized storage network and every computer has to be connected to this specialized storage network; so which means you have to be investing additionally along with this , along with this, along with the storage, the network, the specialized design network which has to be built for managing, for connecting to this storage. So, so because of this additional cost this network, this new network will be built only for server class devices using some optical fiber channel.

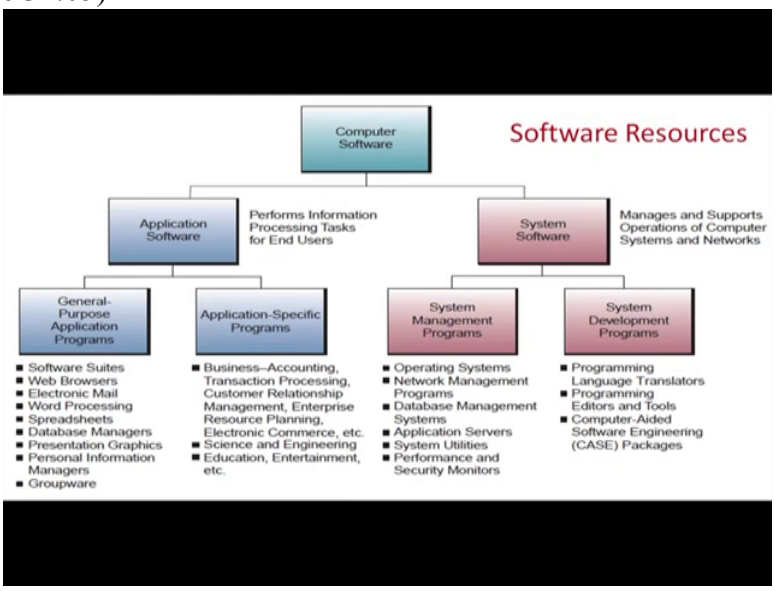
Now this SAN file sharing system, it has, it is again operating system independent, operating system dependent and it does not exist in many operating systems. So if at all you are getting a SAN system you also have to have the right kind of operating system. And there is another restriction, the length of the fiber channel to which these SAN computing devices need to be connected is limited to maximum 10 kilometers. But the benefit, the benefit is the faster communication. In case of NAS, your, you have to communicate with your networking device through, through a, the your ordinary LAN; which means if there is any network congestion etc there will be communication, there will be data loss or similar problems.

(Refer Slide Time 33:09)



This is where we show the major differences, difference between NAS and SAN. Both are actually connected to the network. In case of NAS it is scalable. The data can be transferred over long distances. It is slow and it has the problem of congestion because you are using the same LAN. Then inefficient data backups and recovery because it also depends on DAS devices for this. Then you have, in case of, this is in case of NAS. In case of SAN you have efficient data integrity, backup and recovery and faster, no congestion because it is a specialized network. It is costly and not easily available and data cannot be transferred over a long distance here. Here the maximum limit is 10 kilometers.

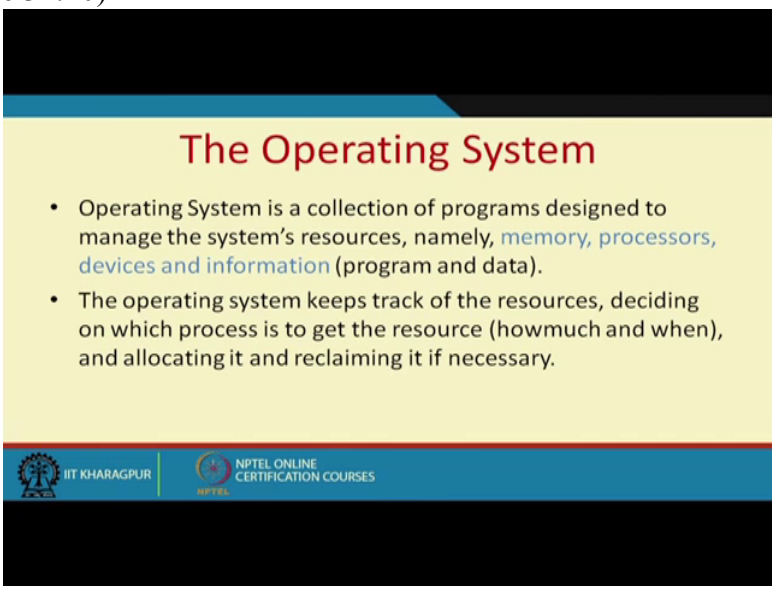
(Refer Slide Time 34:05)



Now we talk about the computer software resources. In fact we will not be spending a lot of time on this. Your computer software is actually, can be either application software and

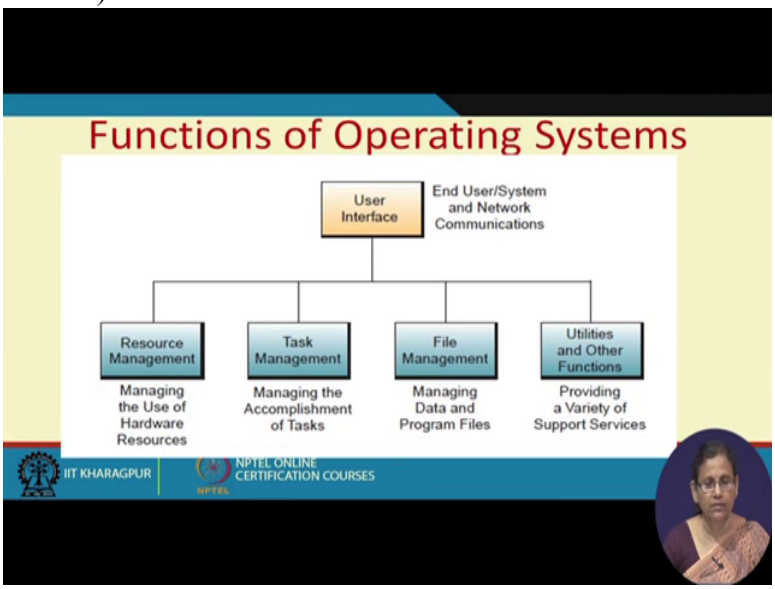
system software, that we have discussed. But one important and all these application layer things, most of them in the application specific programs we have already discussed. Programming language etc we are not going to discuss, only important element here that we need to know little bit more about is your operating

(Refer Slide Time 34:40)



system. Your operating system is a collection of programs designed to manage a system's resources like your memory, processor, device, information etc. Now this operating system keeps track of the resources and shows, and decides when to use and how much to use while sharing the resources among various processes.

(Refer Slide Time 35:07)



These are the typical functions of operating system. It has one user interface through which people can be communicating. This can be a very good user interface like that of Windows or it is some, in case of UNIX systems in case, it may not be a very user-friendly interface but there has to be some interface. Then it has 4 functionalities, resource management, managing and using, managing the use of the hardware resources and task management which is about managing the various tasks given to the operating system, then file management, it is about managing data, program file etc. then providing utility, other utility functions like providing various support services. So with this we finish this lecture.

(Refer Slide Time 36:09)

And from next class onwards we will be talking about the data resources. Thank you very much.