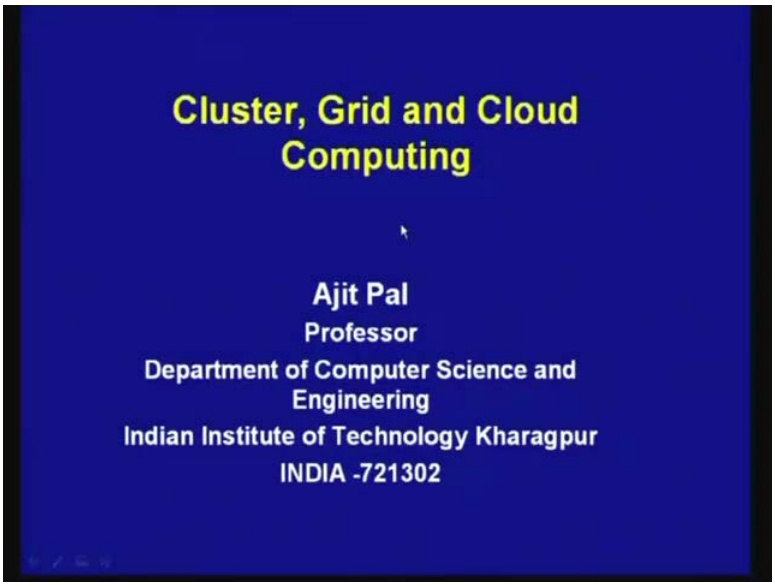
High Performance Computer Architecture
Prof. Ajit Pal
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 41
Cluster, Grid and Cloud Computing

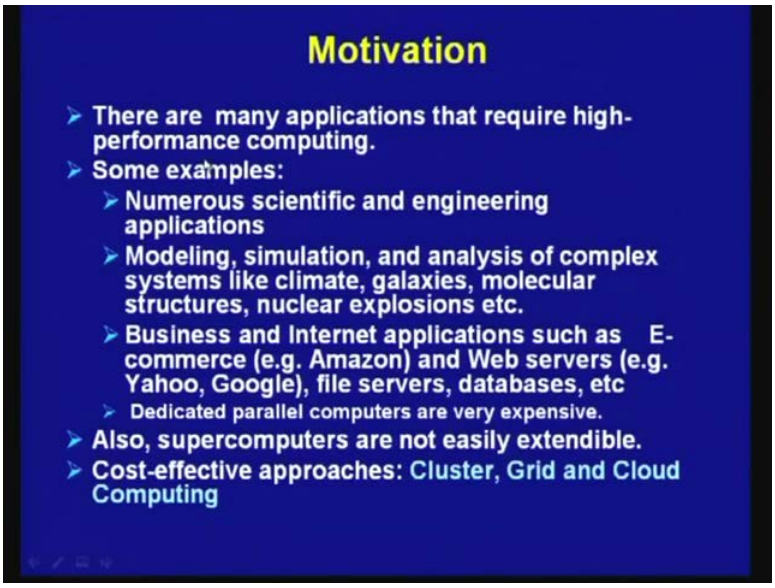
Hello and welcome to this last lecture of this lecture series, on High Performance Computer Architecture.

(Refer Slide Time: 00:42)



In this lecture I shall consider the topic Cluster, Grid and Cloud Computing, so these are the three topics which are little inter related, but definitely they are not same. And I shall consider them one after the other, and as we copper different topics, we shall high light their differences.

(Refer Slide Time: 01:08)



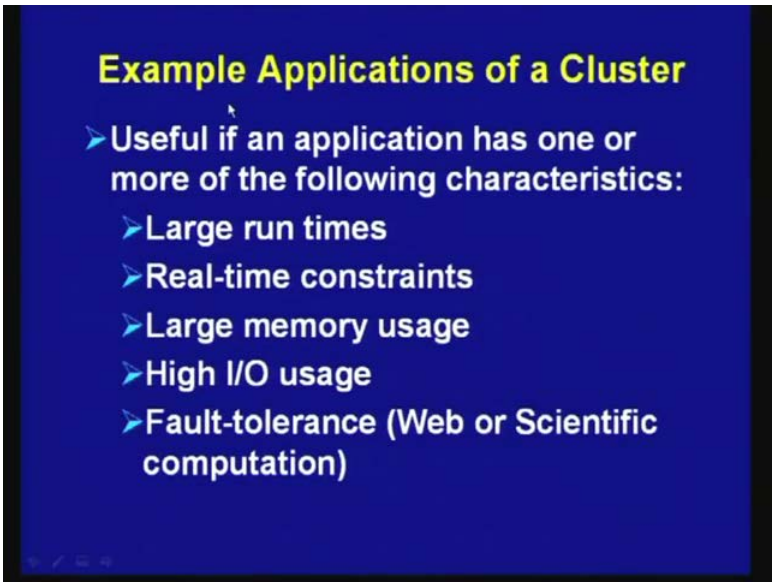
First let as look at the motivation behind cluster computing, we will find there are many applications that require high performance computing. So, numerous applications are now-a-days available are present, which requires high performance computing, and that can be satisfied with the help of cluster computing. And some examples of these high performance computing requirement, I mean which requires high performance computing is given here.

Number 1 is numerous scientific and engineering applications, like modeling, simulation and analysis of complex systems like climate, galaxies, molecular structures, nuclear explosion and whether it is forecasting and so on. And also there are applications which requires high performance computing in business, and internet applications such as E-commerce and web servers, file servers, databases and so on. And for these applications, if you want to have a very dedicated computer, custom made computer with custom software, it is pretty costly. And more over this custom hardware and software does not allow, and not only they are very expensive, they are not extensible.

That means, they can be extended as the requirement goes off, so I mean some kind of super computers are needed to satisfy this high performance computer requirement, but unfortunately they are not extendable. And the cost effective approach is to use cluster, grid and cloud computing for these high performance computing, to satisfy this

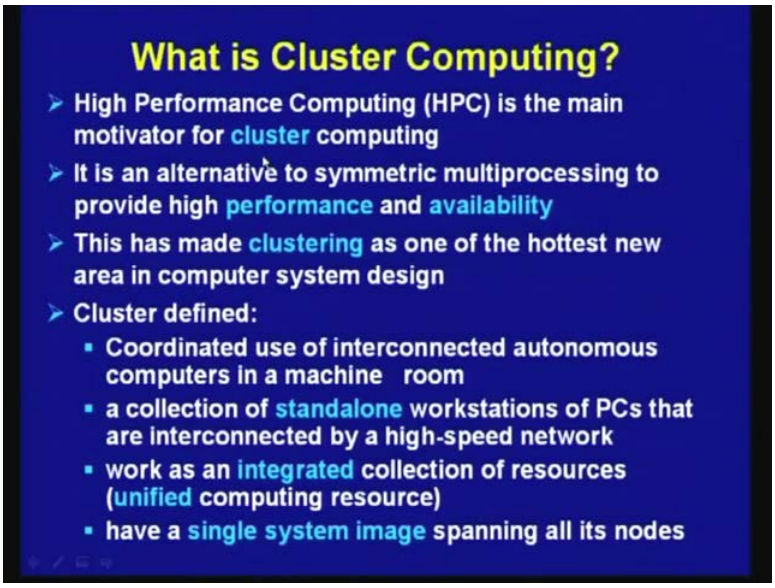
computing requirement, and as you shall see how these are platform in this three situations.

(Refer Slide Time: 03:19)



And as I have already mentioned examples applications of cluster computing, and which has I mean, you may be asking in few situation you will use cluster computing, these are the several situations like, where you require large run times, real time constants that we satisfied, large memory usage is there. Then you require high I O usage, then you require fault tolerance and also you want to have high availability, in such situations we will go for cluster computing.

(Refer Slide Time: 03:57)

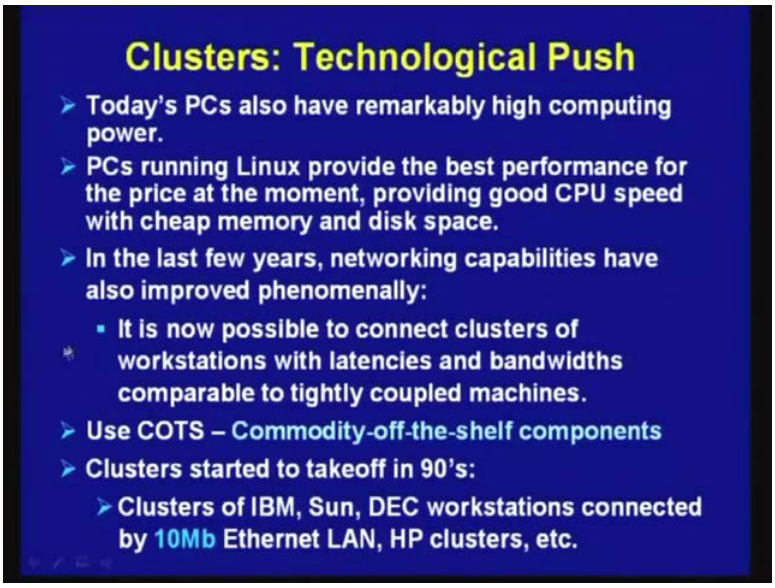


So, as I mentioned high performance computing is the main motivator, and it is an alternative two symmetric multiprocessing to provide high performance and availability. And as a consequence clustering has become a hottest topic in computer system design, you may be asking how do you cluster computing, actually it can defined as coordinated use of interconnected autonomous computers in a machine rooms.

That means, your interested to perform the computing with the help of a interconnected array of computers inside a room, single room, it is not that they are dispersed throughout the country or throughout the large geographical area, but in a machine room. And you notice another term autonomous, by autonomous mean each and every computer is a complete system, can work independently, without the need of others that is you call it cluster computing. So, you can say collection of standalone workstations of PC's that are interconnected by high speed network, may be a local area network, and work as an integrated collection of sources.

So, you find in addition to this the important feature is you have a large number of systems, but at the same time they work in a integrated manner. And they provide a single system image, they have a single system image spanning all it is nodes, so this is a very important concept you will require a special software layer to provide these as usual see.

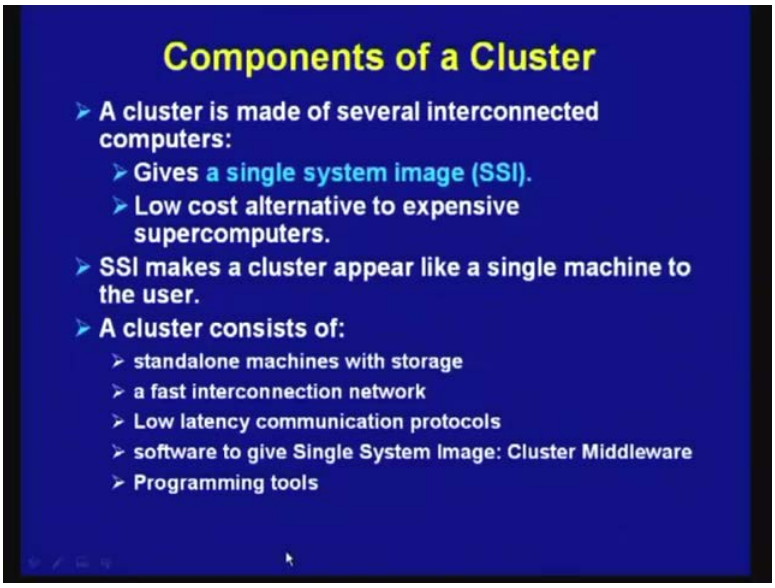
(Refer Slide Time: 06:01)



Actually the cluster computing has become popular, because of two reasons; number one is nowadays you can have personal computers, which are pretty powerful at a very affordable price. Similarly, you can have computer networks like switches, different types of switches, which also available at very low price, so these two is the main motivation. And so that means, it is now possible to connect cluster of workstations with latencies and bandwidths comparable to tightly coupled machines, because of the advancement of your networking technologies.

So, we can build a system by using commodity of the shelf components, what are the commodity of the shelf components, we shall be using standard PC's or workstations and we shall be using standard networking components. So, clusters are started to take off in 90's, clusters IBM, Sun, DEC workstations are connected by Ethernet, 100 Megabyte Ethernet LAN, HP clusters and so on. So, you can see we are using standard local area network to link the different nodes.

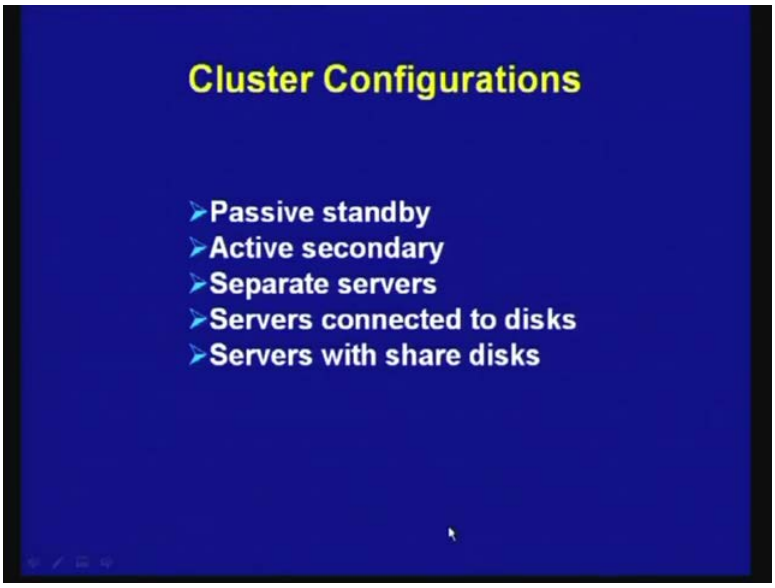
(Refer Slide Time: 07:33)



And as I have already mentioned, although you have got several interconnected computers, but it gives single system is a image. And it gives you high performance, and it is very inexpensive or you can say this is low cost alternative to inexpensive computers. So, a single system image makes a cluster appear like a single machine to the user.

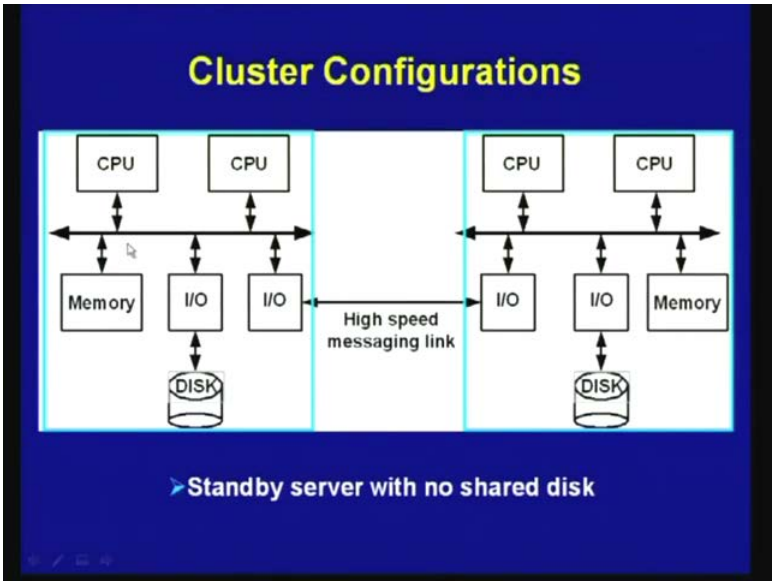
So, you can say if you look at the different components of a cluster, it consists of standalone machines with storage a first interconnection network, high speed LAN, low latency communication protocols. So, not only we will be using hardware, I mean fast interconnected network, you have to use suitable software which will provide you low latency, for communication. And software to give system image that is known as cluster middleware and of course, you will require a host of programming tool to utilize a cluster.

(Refer Slide Time: 08:47)



And there are possible different causal possible configurations, one is passive standby to make it reliable or to make it more available. And active secondary, then you can have separate servers, you can have servers connected to disks, servers with share disk, as it is shown in this diagram.

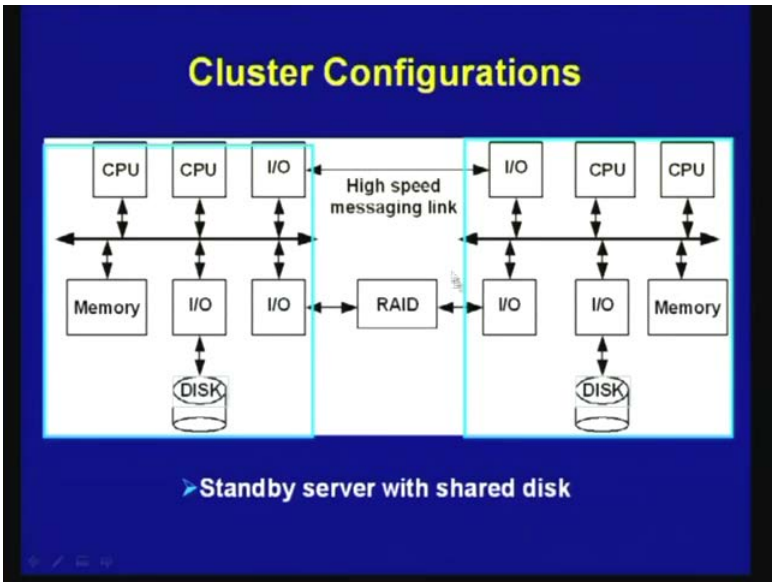
(Refer Slide Time: 09:11)



Here you have got two computers shown here, I mean here you have got one you can say, number of processors connected to a shared bus and this is connected to another system, which is also I mean multiple processor and which shared memory. So, these

two are shared memory multiprocessors, these two are connected together with a high speed messaging link to work as a cluster.

(Refer Slide Time: 09:51)



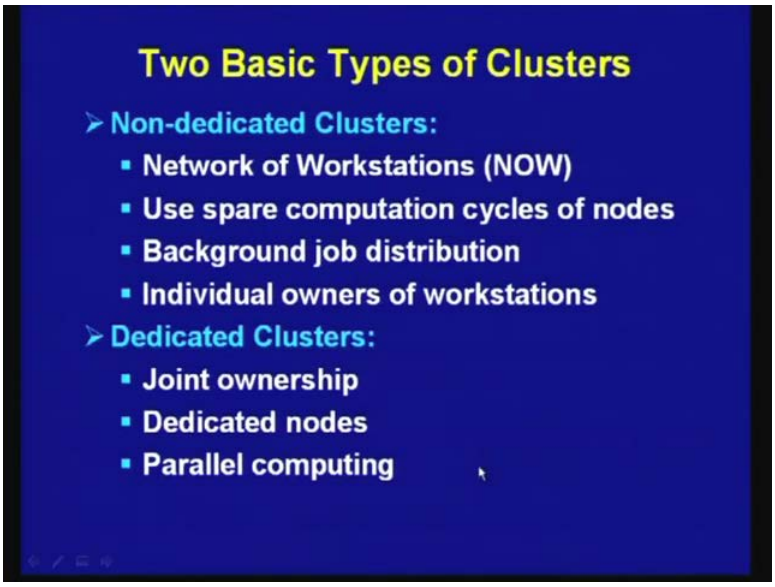
This is another configuration, in which you can have not only high speed messaging link, but you may have a shared hard disk, so in the form of redundant array of in expensive or independent disks. So, here a raid is shared by these two systems through I O buses, so this is stand by server with shared disk.

(Refer Slide Time: 10:15)

And for a cluster you must be having a operating system, and these are the various design issues in the context of clusters, first offal you have to take care of failure management. So, whenever you are behaving a large number of computers, some of them may fail, so it will tolerate failure, then it will be cohere form failure. So, you will require software for operating system will take care of failure management, then you have to do load balancing, the load has to be uniform distributed among all the nodes in a cluster.

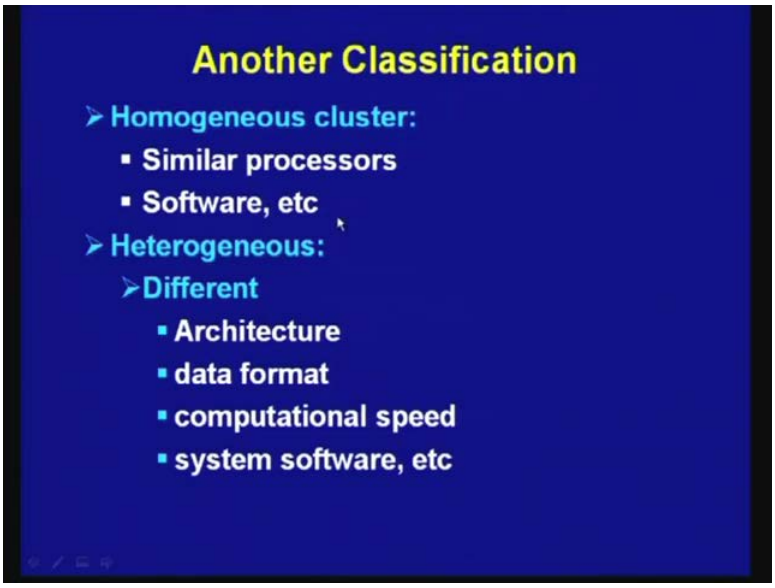
Then you will require parallelizing computation, because you will be doing a kind of parallel processing with the help of the number of commuters. So, you will require parallelizing compiler parallelizing application which is amenable to application, which are amenable to parallelism then parametric computation. So, these are the operating system design issues.

(Refer Slide Time: 11:23)



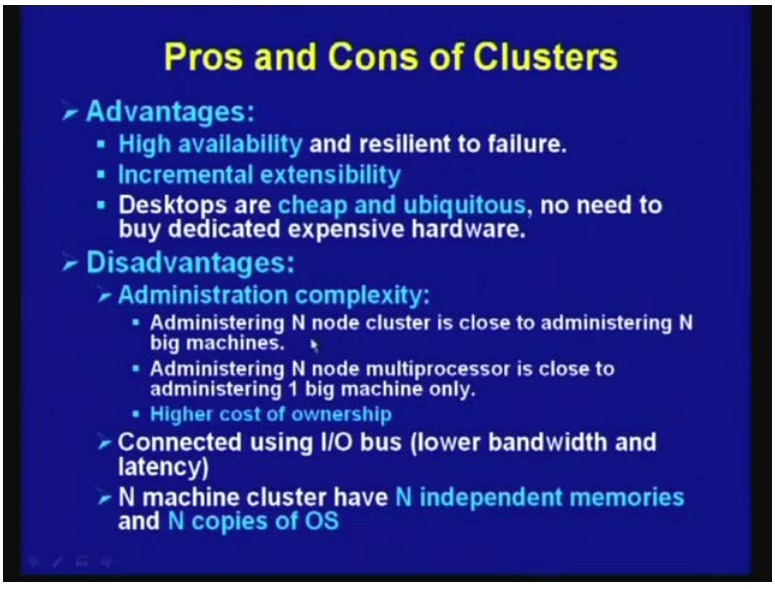
In the context of clusters, then you can have here are the two basic categories of type one can be non dedicated clusters, so they are essentially you may say signature. So, for example, a network of workstations use spare computation cycles of nodes, and background job distribution is done individual owners of workstations. So, this is non dedicated, then you can have dedicated clusters with joint ownership having in dedicated nodes, and which will also allow parallel computing.

(Refer Slide Time: 12:04)



Then you have got, you can have homogenous cluster another way of classifying is homogenous and heterogeneous, so homogeneous cluster will have similar processors. So, all the nodes will be identical or similar software operating system, and software present in all the components, all the nodes in cluster. Alternatively you can have heterogeneous system with different architecture, data format, computational speed, system software that means, operating system and so on.

(Refer Slide Time: 12:35)



So, let us see what are the advantages and disadvantages of the clusters obviously, main advantages is gives you high availability, and resilient failures. That is since, you have got large number of computers, if one or two fails still it will continue to function and as a consequence it gives you high availability. And then second important feature is incremental extensibility, so as your requirement keeps on increasing with time, you keep on adding nodes in a cluster. So, you can increase the number of nodes in a cluster, in a incremental manner.

And desk tops are cheap and ubiquitous, so this is another important advantage, so you are able to provide high performance computing at a very low cost, no need to by dedicated expensive hardware. So, you can use commodity of self type of desk tops for your application for developing a cluster, so these are the advantageous. Of course, you have got several disadvantages, number one is administration complexity, so since you have got you are administering N node cluster, I mean administering N node cluster is close to administering N big machines.

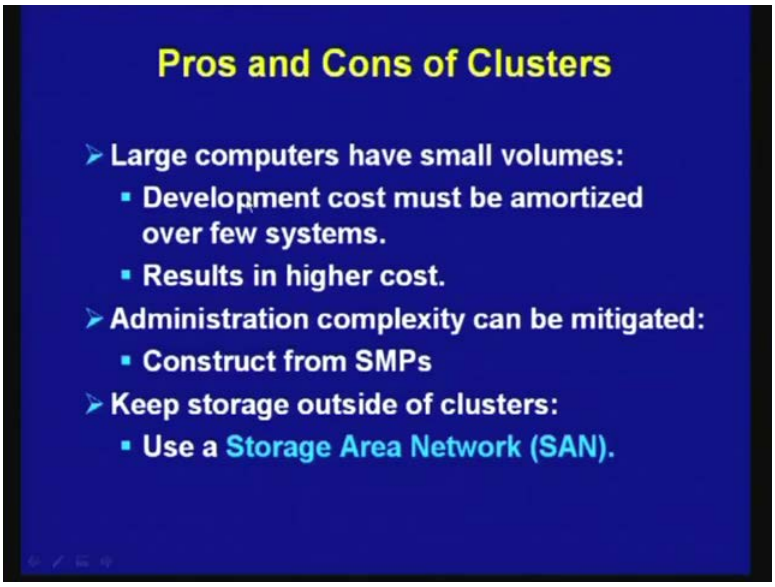
The reason for that is each machine is autonomous with hardware, operating system, application software and so on, so as a consequence each of these nodes are be administered independently and separately. So, administering N node multiprocessors is close to administering one big machine, so whenever you have got multiprocessor systems, shared memory multiprocessor systems, they are you can have only it may be considered as a one big machine.

But, that is not true in the context of clusters, so and also it gives you higher cost of ownership and second important disadvantage is, these computers are connected through I O bus. So, in case of we have seen shared memory multiprocessor system or massively parallel processing system, the processors are connected through memory box. So, memory box has got higher bandwidth and smaller latency, on the other hand whenever they are connected using I O bus, the bandwidth is smaller and latency is larger.

So, lower bandwidth and higher latency you can say, so higher part is missing here, so it will higher latency, so this is the disadvantage. And another important disadvantage is N machine cluster have N independent memories and N copies of operating systems as I mentioned earlier. So, since each is autonomous will be having independent memories,

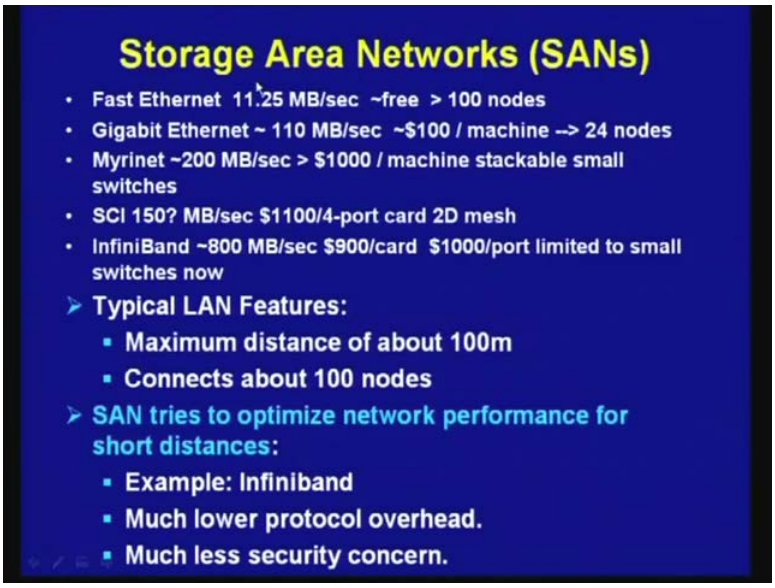
so that they can work independently and N copies of operating system, so because each can again operate independently without need of others.

(Refer Slide Time: 15:50)



And here is another advantage large computers of small volumes, so whenever you are building a large computers, then cost has to be amortize over few systems. So, these results in higher cost, on the other hand in case of clustering, you have got large number of low cost system as a result the cost is much smaller. So, administrated complexity can be mitigated, construction of shared memory multiprocessors and keeps storage outside the clusters. So, you can keep the storage outside the clusters, as it happens in case of storage area network.

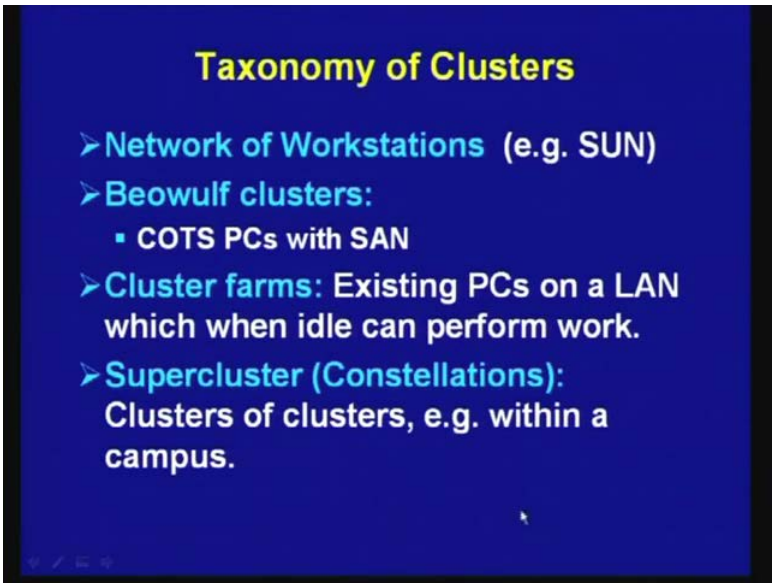
(Refer Slide Time: 16:41)



So, these are gives you some components which are use like, the local area network that is been use, you can use high fast Ethernet or Gigabit Ethernet or Myrinet, so these ate type networks that can be used for connecting or InfiniBand, which gives you 100 Megabytes per second. So, you can see these are the rates 11.2 high Megabits per second or 110 Megabits per second or 200 Megabits per second or 800 Megabits per second, so this speed is quite high and distance maximum.

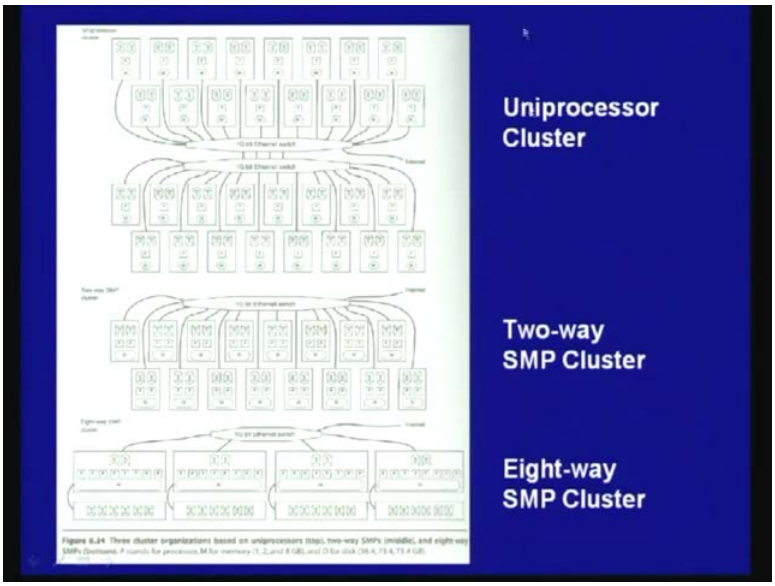
Of course, distance is limited you can it has to be within 100 meters that is why I said that it has to be within a room, but it did not be really within a room, so 100 meter where is quite long distance and you can connect about 100 nodes. So, storage area network tries to optimize network of performance for short distances for example, as it happens in case of InfiniBand. So, you have much slower protocol over head and much less security concerned, as it happens whenever you do the communication through internet.

(Refer Slide Time: 17:55)



So, let us have a look at the taxonomy of a clusters, one is network of workstations example is Beowulf clusters, where it use COTS PC's that we commonly, I mean commodity of the self PC's with storage network. And you can have cluster forms existing PC's on a LAN, which can hardly can perform work or super cluster or constellation, where you can have cluster of cluster that is within a campus. So, these are possible ways in which you can middle clusters.

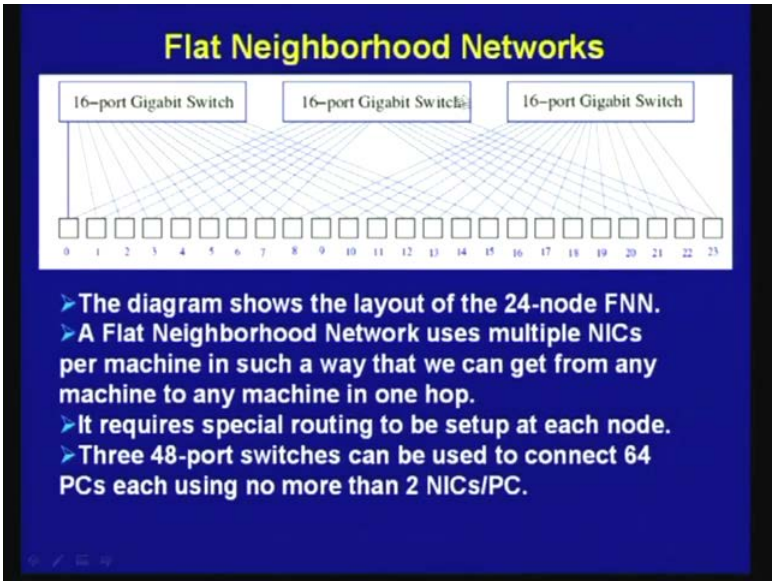
(Refer Slide Time: 18:44)



So, these are some of the examples, the club diagram shows uniprocessor clusters, each processor here you have got a single processor with it is own dedicated memory, and I O. So, you have got a large number of such processors connected through a Ethernet switch, 1 GP Ethernet switch, here you have two way symmetric multiprocessor cluster, because each node is having two processors with a shared memory and I O. And these are connected through 1 GP Ethernet switch and the third diagram that is shown is eight way SMP cluster.

Here you have got eight processors in each node with a shared memory and I O and switches connected with the help of again 1 Gigabit Ethernet switch and of course, may be connected to internet. So, these are the three different types of cluster shown here, they are shown in the processor nodes to a SMP nodes, are eight way SMP nodes.

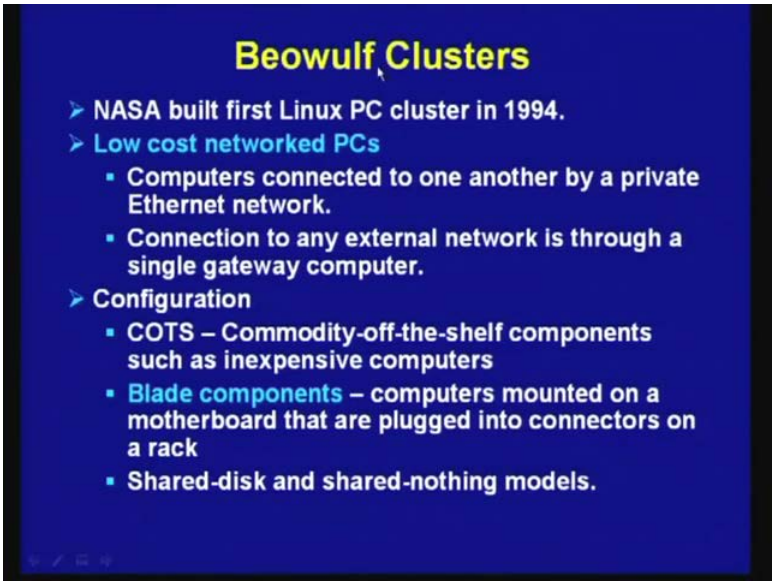
(Refer Slide Time: 19:57)



And this is how the connections can be done flat neighborhood networks, this is how the interconnection can be made and this particular topology is known as Ethernet or flat neighborhood networks. So, this diagram shows the layout of 25 node FNN, a Flat Neighborhood Networks, here as you can see you have got 24 nodes, 1 2 up to 23, 0 to 23, 24 nodes each having processors memory and I O devices. And these are connected with the help of 16 port Gigabit switch, so 16 port Gigabit switch, but one point you should notice is that each node is connected two switches.

So, you have got multiple network interface card, NIC stands for Network Interface Card that means, each node is having multiple in this particular case two network interface card, each network interface card each network interface card is connecting to one of these switches, so two NIC's connecting to two switches. So, in this way the communication between any two node can be done using a single hop, so it requires special routing to the setup at each node 348 port switches can be used to connect 64 PC's each using no more 2 NIC's per PC's. So, this is a very interesting topology and it can be used to realize clusters.

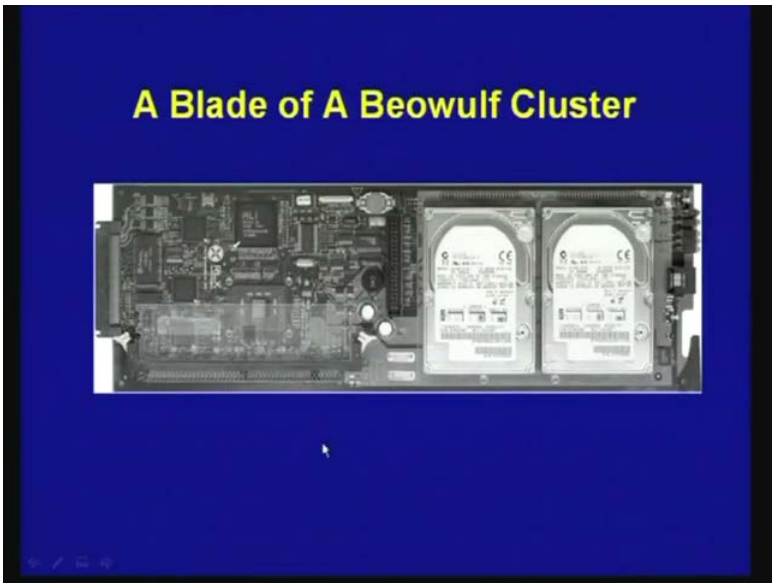
(Refer Slide Time: 21:53)



This is an example Beowulf cluster, NASA build first Linux PC cluster in 1994, so it is a low cost network PC's computers connected to one another by a private Ethernet network. So, connection to any external network is through a single gate way computer, and configuration is you are having commodity of self components, such as inexpensive computers. And blade components are use that means, the computers mounted on a mother board that are plugged into connectors on rack.

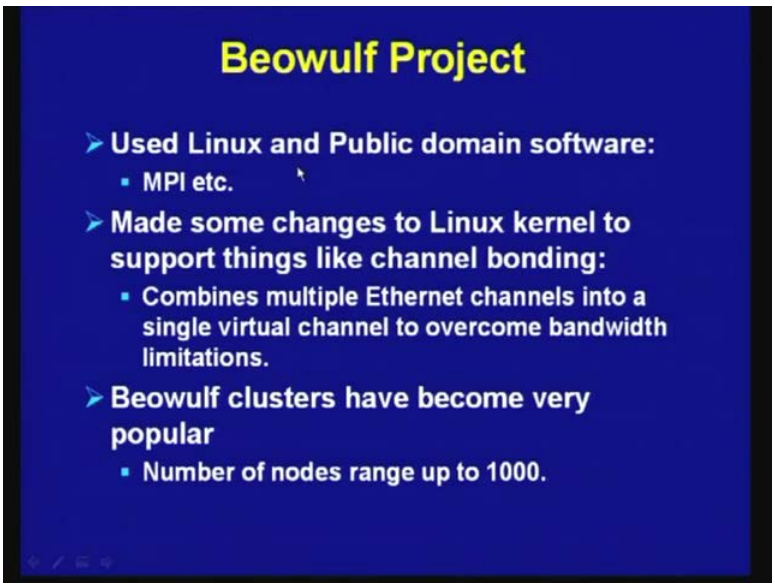
So, you will be using a rack on the back plane, there are connectors and each of the blades can be pushed to connect to the system as I shall show you, and shared this can shared nothing model is possible in whenever you go for this type of cluster.

(Refer Slide Time: 22:50)



This shows a single Beowulf cluster, as you can see this is the connector which can be connected to the back plane. And here it shows two processor, this is one processor, this is another processor to two processing module, which are present in a single blade.

(Refer Slide Time: 23:14)

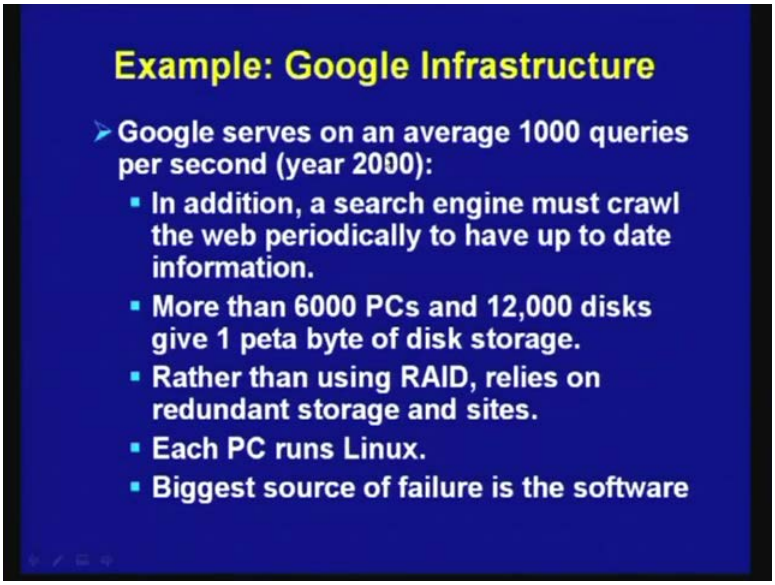


And this Beowulf project of NASA used Linux and public domain software, so they did not go for any custom software, so custom software is not being used, standard commodity of the software has been used. So, made some changes to Linux kernel to support things like channel bonding, so you have seen multiple Ethernet channels are

used to connect, but to connecting to a single virtual channel to overcome bandwidth limitations. We know that the Ethernet has a limited bandwidth, like 1 Megabit per second or 10 Megabits per second, whenever you go for Gigabit Ethernet.

But, if you want higher bandwidth, you can use a special type of technique and this is what is being done known as, with the help of this special technique you can combine multiple Ethernet channels into a single virtual channel. So, you can have more than 1 Gigabit bandwidth, so this will overcome bandwidth limitation. And Beowulf clusters have become very popular, and you can have 1000 nodes in a single system.

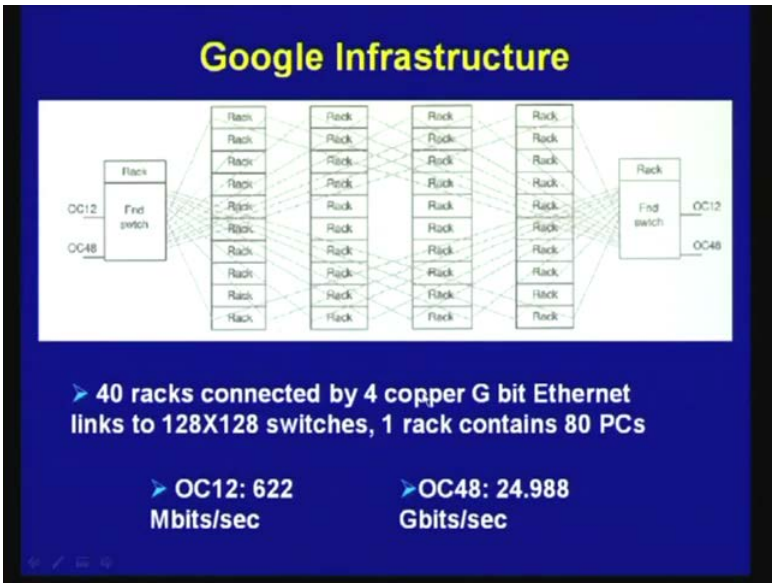
(Refer Slide Time: 24:40)



Another example is Google infrastructure, Google serves on an average 1000 queries per second, all of your familiar with I mean use of, when your using the cyber provided by Google or for E-mail that is G-mail and all those things, so that is service to with the help of this infrastructure. So, in addition to this the serving the queries a search engine must crawl the web periodically to have up to date information. So, this is used for web servers more than 6000 PC's and 12000 disks gives 1 peta byte of disk storage, rather than using RAID this Google infrastructure relies and redundant storage and sites, so each pc runs Linux.

And this is only the biggest source of failure is the software, in case of this infrastructure, because you have got lot of redundant, I mean PC's computers and other things, so only biggest source of failure is software.

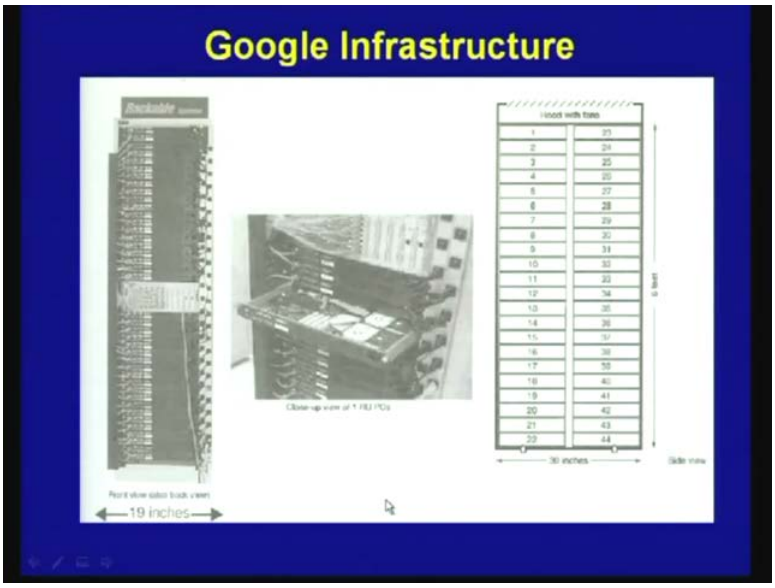
(Refer Slide Time: 26:01)



So, this diagram shows Google infrastructure, you can see here different rack shown 10 racks in here, 10 racks here, we have got 40 racks connected by 4 copper Gigabit. And that links 128 into 128 switches, so we have 128 into 128 switches and 1 rack contains 80 PC's. So, each rack is containing 80 PC's and you have got 40 such racks, so you can imagine the total number of such some PC's have represent here.

And you are using that OC 12 and it is OC 45 that gives you 622 Megabits per second or 24988.988 Gigabit per second that I mean, that is the rate at switch you can communicate with this end switch. So, end switch end gives you faster access to the entire infrastructure, and you have got two such ends switches at both ends.

(Refer Slide Time: 27:22)

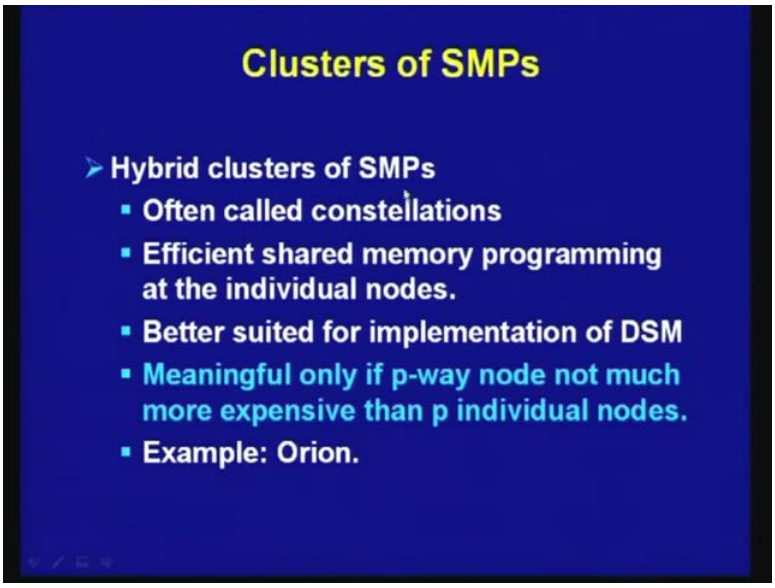


So, this is the diagram which shows the way, this is single rack and close of view of 1 rack PC, so this diagram shows you Google infrastructure.

(Refer Slide Time: 27:42)

There is another view of the cluster, that is implemented with a help of Google infrastructure.

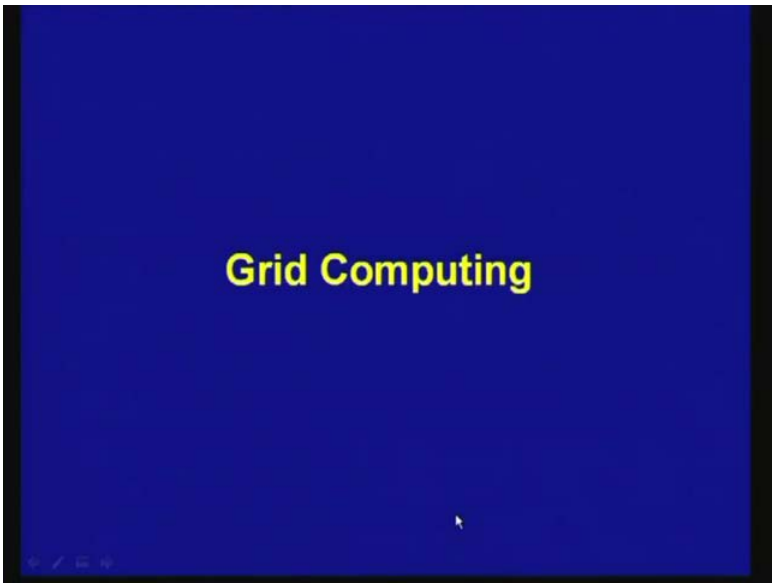
(Refer Slide Time: 27:51)



Then you can have cluster of shared memory multiprocessors, so highbred cluster of shared memory multiprocessor can be used, this is often called constellation. So, this is efficient shared memory programming at the individual node is possible, because each is having a shared memory multiprocessor. So, this is better suited for implementation of distributed shared memory multiprocessor, and this is meaningful only p-way node not much expensive than p individual nodes.

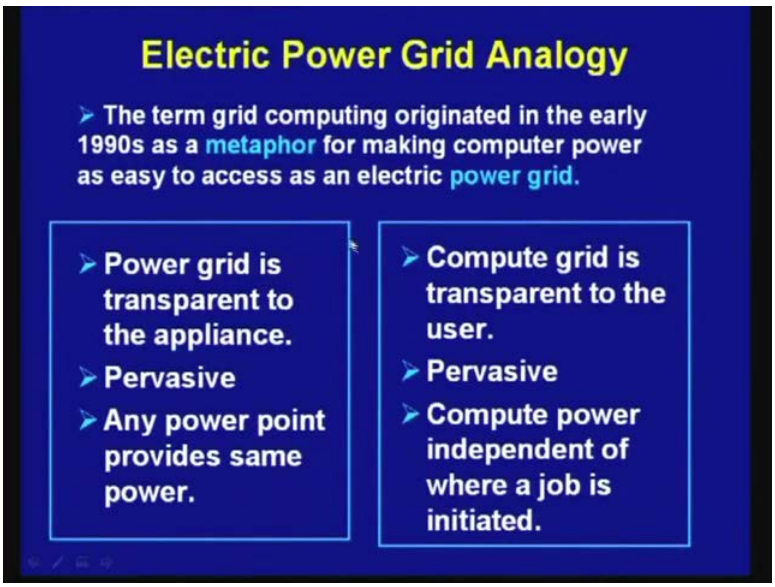
So, this is what has been observe that, if you implement a p-way node that means, each node is having p computers that is cheaper than to have p individual nodes, and then this is meaningful and that is what is being done and example is Orion system. So, we have discussed about one very important topic that is your cluster computing, and how clusters have been implemented and various topologies and at the end we have shown some example, now we shall focus or attention to another way of computing.

(Refer Slide Time: 29:08)



Grid computing.

(Refer Slide Time: 29:15)



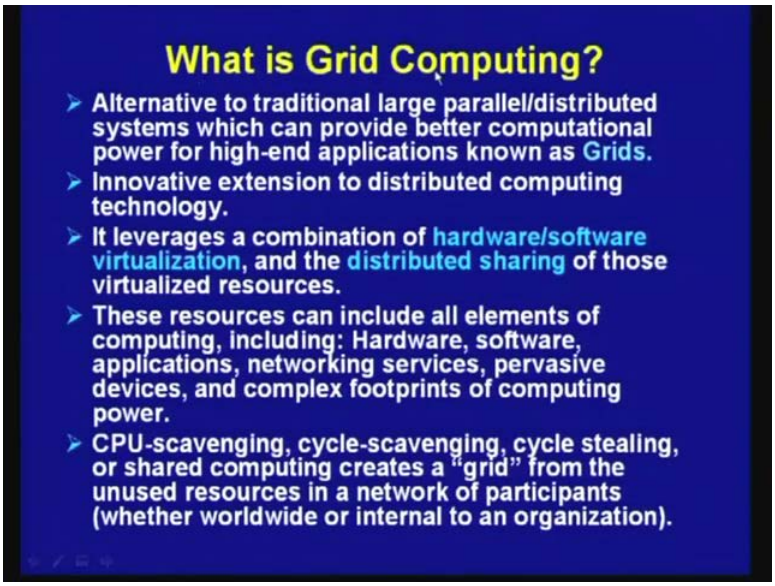
So, why the name grid computing, so this the term grid computing originated in the early 1990's as a metaphor for making computer power has easy to access, as an electric power grid, all of are your using electric power. So, as there is a that electric line is connected to each and every house, and you can access it without much a botheration, so we know that each plug point, we provide you to 20 volt or 120 volt a depending on where you are. And that is the main feature, so power grid is transparent to the appliance, so

whether you will be running a computer or will be running some other appliance, it does not matter you get the power.

And you can run different types of systems, without bothering where from the power is coming and how the power is available and it is all pervasive. So, it is available everywhere and any power point provide same power that means, you get 220 volt at each every point. So, that is power availability is same at different points, to that is what is the main feature of grid computing, so availability is uniform from different places. Similarly, just like power grid, in compute grid is transfer rent to the user, so this compute grid is transparent to the user, although it is distributed throughout the country.

That mean, it is access through wide area network and it is pervasive and computer power is independent of fair a job is initiated. So, you can initiate a job from any fair and you will get the same computing power for performing an application.

(Refer Slide Time: 31:26)

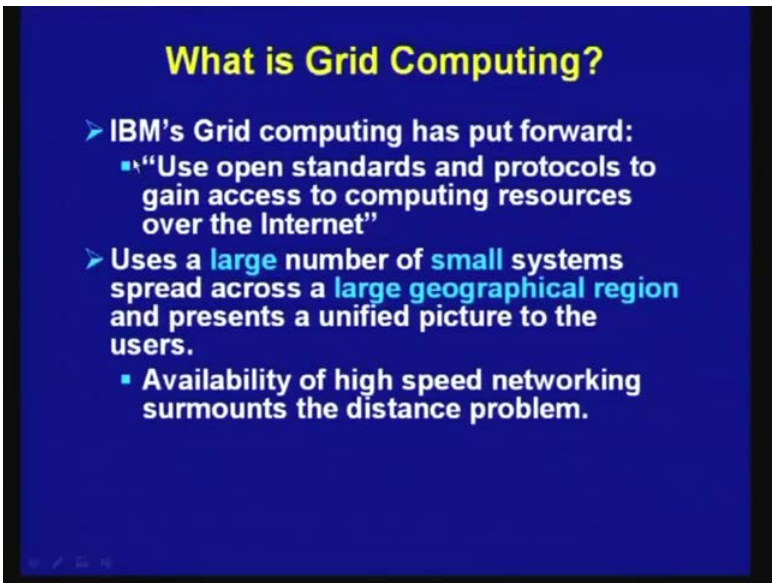


So, this question a sadly arises, what do you mean by grid computing, so this is alternative to traditional large parallel or distribute systems, which can provide better computational power for high and application, this is known as grid. And this is an innovative extension of distributed computing technology, so grid computing is a kind of distributed computing. But, it is not the way traditional distributing computing is done, so it leverages a combination of hardware and software, virtualization, and distributed sharing of those virtual resources, these are the two basic concept.

First up all hardware and software virtualization, second is distributed sharing of this virtualized resources, and these resources can include all elements of computing. So, like hardware, software, applications, networking services, pervasive devices and complex footprints of computing power. So, you can see, you can have different types of a different elements which are virtualized an available, so it can be actually sometimes grid computing is also referred as CPU scavenging or cycle scavenging or cycle stealing or shared computing, creates a grid from unused resources or in a network of participants.

So, what is happening here, you have got large number of participants and all of them performing their own job however, each of them may have some additional or extra computing power, and which is been utilized in grid computing by others through the internet. So, where worldwide or internal tune organizational, so this grid computing can be done through, I mean across the world or internal tunnel to organization.

(Refer Slide Time: 33:43)

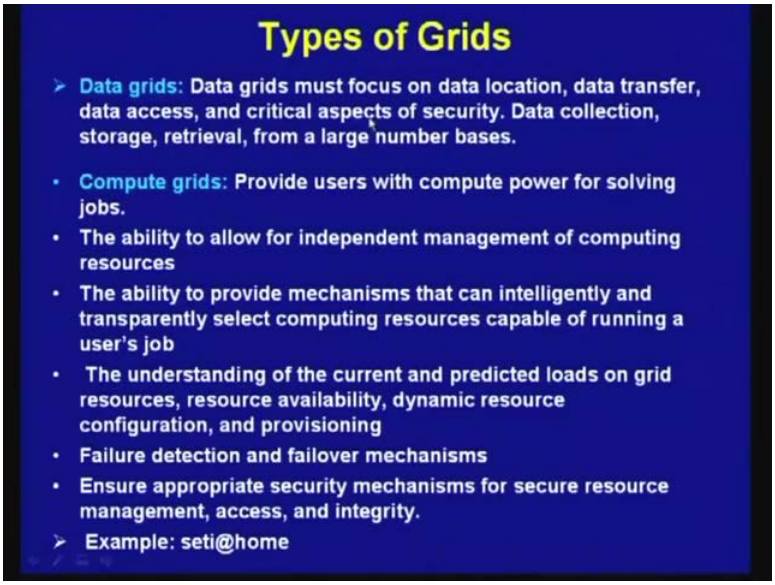


So, IBM's grid computing has put forward that actually basic idea was use open standards, and protocols to be an access to computing resources over the internet. So, you are gaining a access of computing resources over the internet, and you are using large number of small systems spread across a large geographical area region, and it presents unified picture to the users. So, here also it is giving some kind of unified

picture to the user, so you are not really concisely accessing, each and every different computer, so you are accessing a grid. So, perform you are getting is transparent to you.

And this availability of high speed networking surmounts the distance problem, you may be asking how you are able to get high performance through wide area network. The reason for that is, because of the advisement of wide area network technology, now you can have high bandwidth, the high bandwidth of access through internet. And that how this availability of high speed networking is possible, and it is surmounts the distance problem.

(Refer Slide Time: 35:15)



So, the grid's can be divided into several types, one is your data grid another is your compute grid. So, data grids must focus on data location, data transfer, data access and critical aspects of security, because you are handling data, that data collection, storage, retrieval from a large number of bases. So, this is the function of a data grid similarly, you can have compute grids, it provides user with compute power for solving jobs. So, the ability to allow for independent management of computing resources, the ability to provide mechanisms that can intelligently and transparently selecting computing, resources.

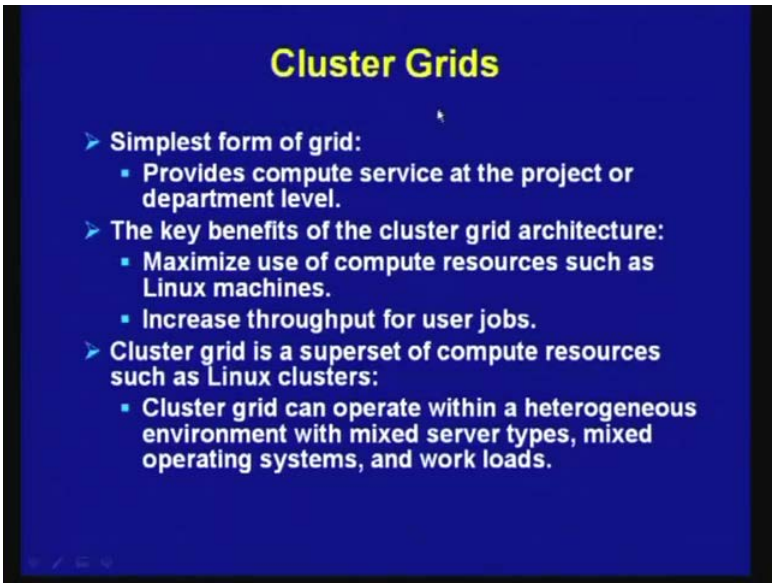
Capable of running a users job. the understanding of the current and predicted loads on grid resources, resource availability, dynamic resource configuration and provisioning. So, these are the features than failure detection and failure mechanisms, so the essential

appropriate security mechanism for secure resource management access and integrity, and example is seti@home, this is an example of a grid computer.

(Refer Slide Time: 36:35)

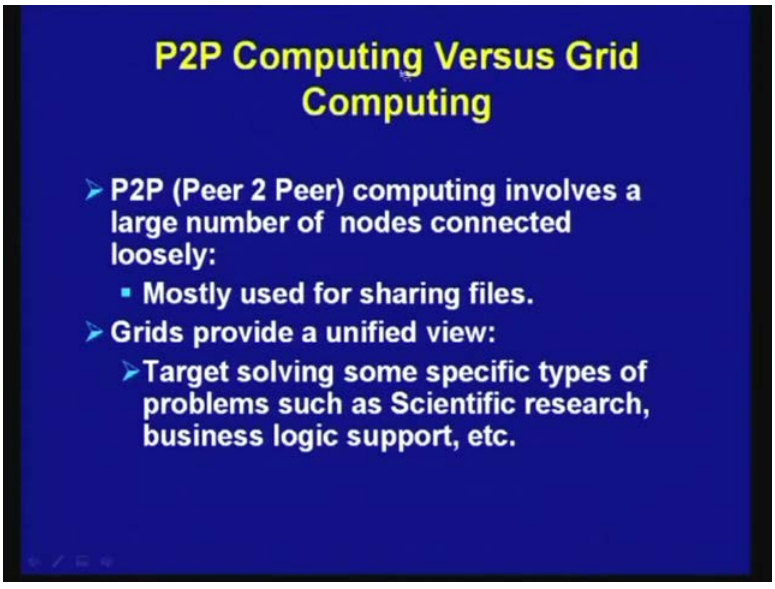
And you can have three different types of grid environments, one is global grids, second is enterprise grids, third is cluster grids. So, global grids are collection of enterprise and cluster grids available across the globe, an enterprise grids are multiple projects of departments here, resources within a enterprise your campus, so this is enterprise grids. So, it does not do not at the security and global policy management issues, because since it is restricted within enterprise you do not really require, it is not necessary to address security. And global policy management issues, which is required in place of global grids.

(Refer Slide Time: 37:23)



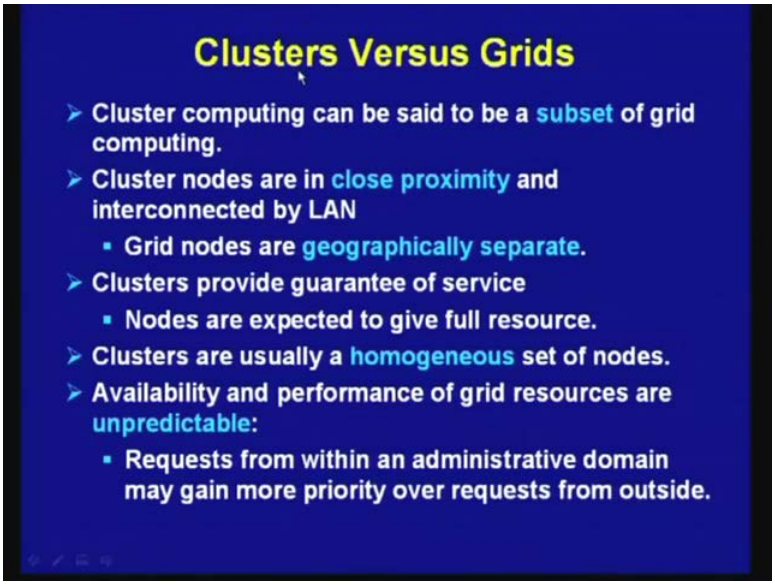
And third type is cluster grids, it is a simplest form of grid, it provides compute service at the project or department level. The key benefits of cluster grid architectures are number one is maximize use of computer resources, such as Linux machines and increase throughput for user job. So, these are the two key benefits, that you achieve with the help of cluster grids and cluster grid is superset of compute resources, such as Linux clusters. So, cluster grids can operate with a heterogeneous environment with mixed server types, mixed operating systems and workloads, so this is another very important feature, you can have heterogeneous environment.

(Refer Slide Time: 38:13)



So, this is a comparison between P2P computing and grid computing, all of you are familiar with Peer 2 Peer computing, which involves a large number of nodes connected loosely, and this P2P computing is primarily used for sharing files. So, you can distribute files over a large number of nodes, then for efficient sharing you distributes, then you access from different nodes, that is the basic idea of Peer 2 Peer computing. On the other hand, grids provide unified view, so target solving some specific types of problems such as scientific research, business logic support etcetera, so it is not restricted to all if you have file sharing.

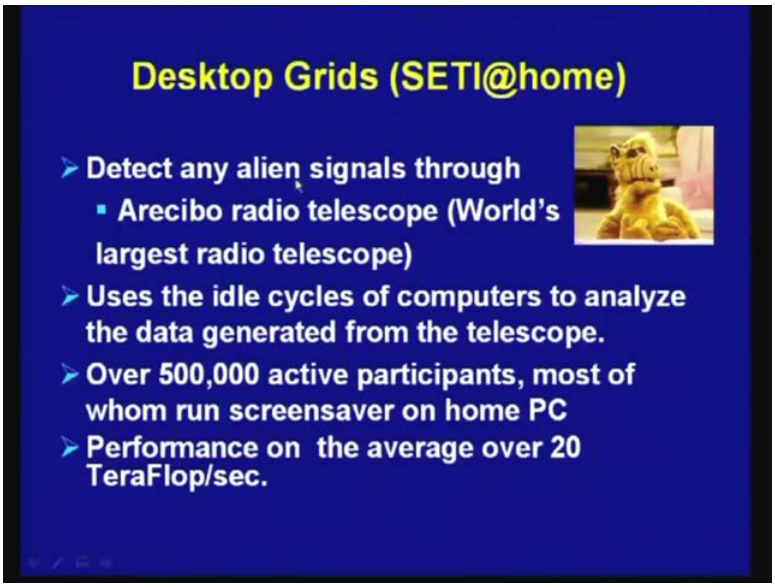
(Refer Slide Time: 39:05)



So, here is a comparison between cluster computing versus grid computing, so as I mentioned cluster computing can be said to have a sub set of grid computing. Because, cluster nodes are in close proximity, and interconnected by LAN as I have mentioned, and grid nodes are geographically separate and connected through wide area network or WAN. So, clusters provide guaranty of service, nodes are expected to the full resource on the other hand, clusters are usually heterogeneous site of clusters.

So, since it is distributed throughout the network availability and performance of grid resource are unpredictable; and request from within an administrative domine may gain more priority of over request from outside. So, this is another feature that has led to I mean, that differences in availability of different systems, availability and performance of heat resources are unpredictable because of many reasons and this is one of them.

(Refer Slide Time: 40:22)



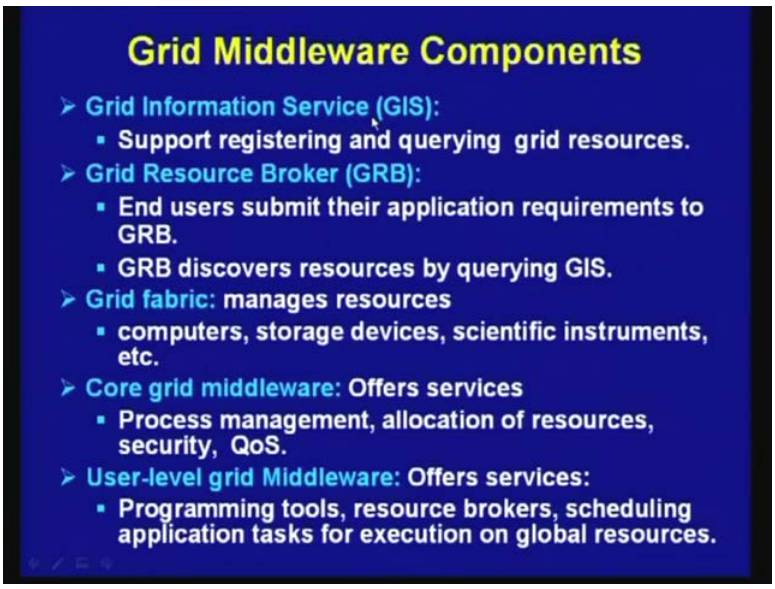
So, this is that example SETI at home that was develop for to detect, and alien signals through Arecibo radio telescope, world is largest telescope, so this used the idles cycles of computers to analyze the data, generated from the telescope. So, over 500000 active participant most of whom run screensaver on home PC and so from these active participants the computing power is used to analyze the data generated from the telescope.

And performance on the average over 20 teraflop per second, so you can see, you can have you can achieve massive performance, because of large number of participants, although each of them is giving only affection of their computing power.

(Refer Slide Time: 41:27)

Now, the big question is at some level and all application share common needs, how it is being done, so how to find resources, how acquire resources, how to locate and move data, how to start and monitor computation, and how to manage it all securely and conveniently. So, for that purpose the software that is being used is known as grid middleware, so the single software infrastructure supports all the above features.

(Refer Slide Time: 42:06)



So, these are the different middleware components, number one is grid information service GIS, and support this supports registering and querying grid resources, grid

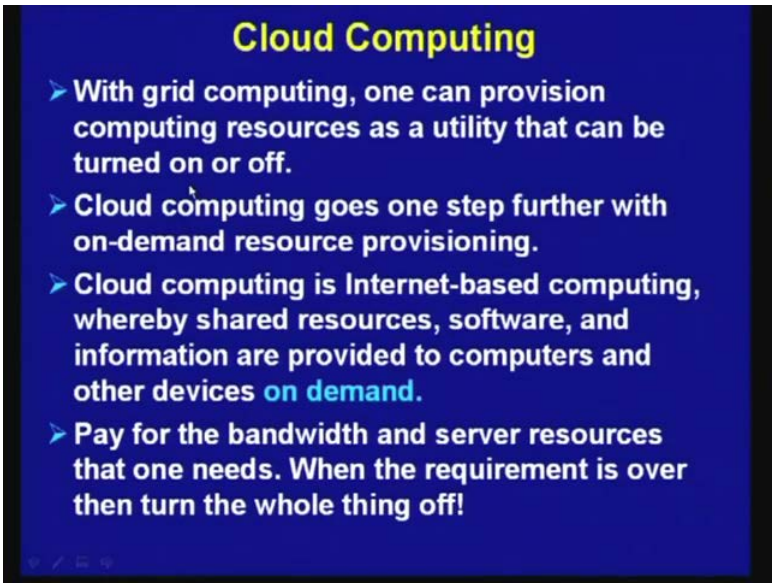
resource broker end user submit their application requirements to this grid resource broker. Then grid resource broker discovers resources by querying the GIS, GIS means this grid information service, then you have got grid fabric which manages resources like computers, storage devices, scientific instruments etcetera.

Then you have got core grid middleware, this offer services like process management, allocation of resources, security and quality of service and so on. And user level grid middleware, this so far services like programming tools, resource broker, scheduling application tasks for execution on global resources. So, you can see this grid computing is feasible, over large geographical area with the help of these grid middleware components.

(Refer Slide Time: 43:23)

Now, we have come to the last topic that is your cloud computing.

(Refer Slide Time: 43:32)



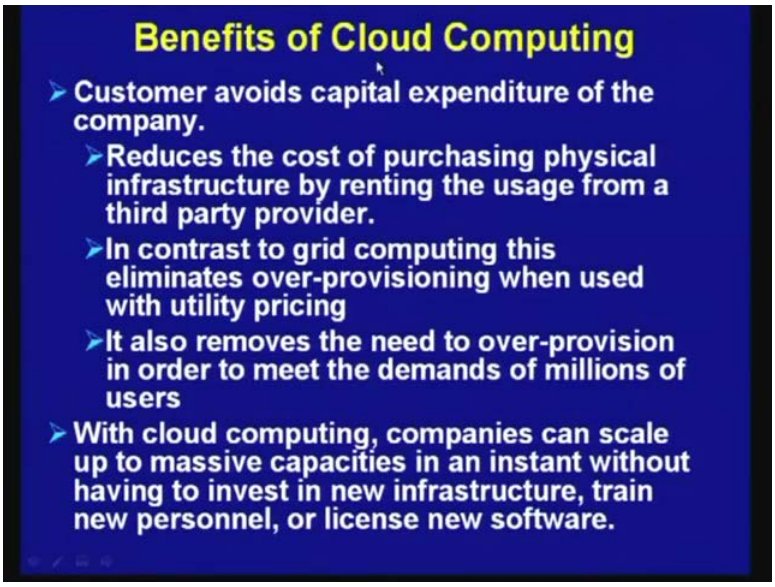
Cloud computing is also a kind of distributed computing, I mean somewhat similar to grid computing, in case of grid computing we have seen with grid computing, one can provision computing resources as a utility that can be turned on and off. So, here in grid computing it is done in this way, you can provision a resource and after the provisioning is done, you may utilize the entire resource or may utilize it part partially. And then if you do not need it, then you can turned it off that means a particular resource provisioning can be done, and you turn it on or off.

But, whenever an utility is being used, it is not guaranteed that you are utilizing the entire resource and this cloud computing goes one step further with on-demand resource, provisioning. So, here we are not really I mean a particular resource is turned on and then, it does not allow instead of turning it on and using it partially or not in this case. A concept called resource provisioning is done, so resource provisioning means you need, whenever you need and as much you need it you will use that.

So, this is the on-demand resource provisioning in contrast to on off type resource, process provisioning is done in grid computing. So, cloud computing is internet based computing, where by shared resources, software and information are provided to computers and other devices on-demand. And one big advantage of this on-demand resource provisioning is that, you pay for the bandwidth and server resources that one

need. And when the requirement is over, then the turn the whole thing off, so you have to paying only for that part you are using.

(Refer Slide Time: 45:53)



And here are the benefits of cloud computing, customers avoids capital expenditure of the company that means, here what you are doing you are not really deploying costly infrastructure, so you are using third party infrastructure to perform you job, so this is one very important concept. So, this reduces the cost of purchasing, physical infrastructure by renting users from a third party provider, so you are only renting a part of the resources provided by a third party.

So, instead of in contrast to gird computing this eliminates over provisioning, and used with utility pricing that means, instead of as I mentioned whenever in grid computing it may lead to over provisioning. But, in case of you cloud computing, over provisioning is not done and as a consequence you pay only a very small part and when used with utility pricing. That means, you are using at an utility and as much as you used, you pay for that and it also removes the need to over provision, in order to meet the demands of millions of user.

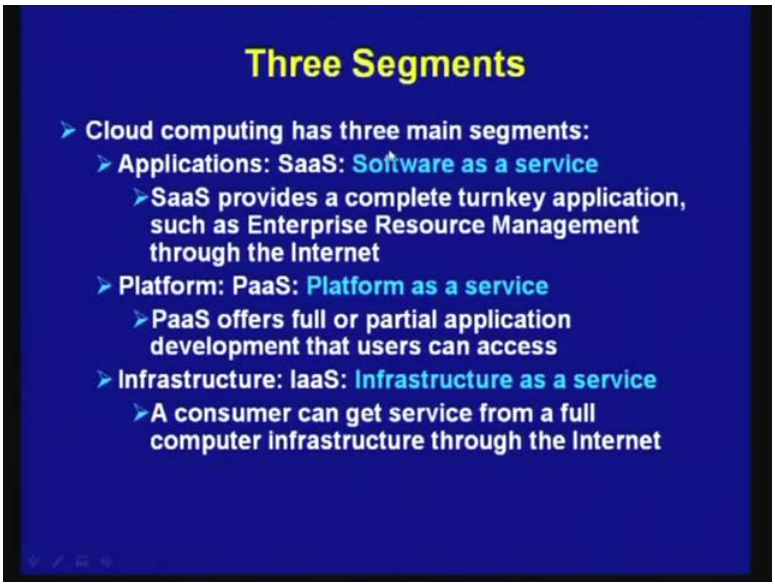
You see in not only you are paying less, but it also helping others this is what understand, so you are taking or you are consuming only as much you can, so the remaining part is available for others. So, in this way grid computing can offer to grid service to a large number of users, and with grid computing companies can scale up to

massive capacities in on instants without having to invest in new infrastructure, train a new personal or license new software. So, you can see, a company whenever you need massive computing, it can scale it up by hiring resources from a third party.

And as soon as the this high performance computing is over to simply turn it off, similarly you can use massive storage, whenever you need it and whenever you do not need it you simply turn it off. And also you can use a special software, there is no need for to license a new software, so you are not doing the licensing, the license is not your name, license is in the name of the service provider, you are simply using it and you are paying for your uses.

So, without having license of the software, without purchasing the software, without purchasing the hardware you are able to use them, to your advantage, to your benefit this is the benefit of cloud computing.

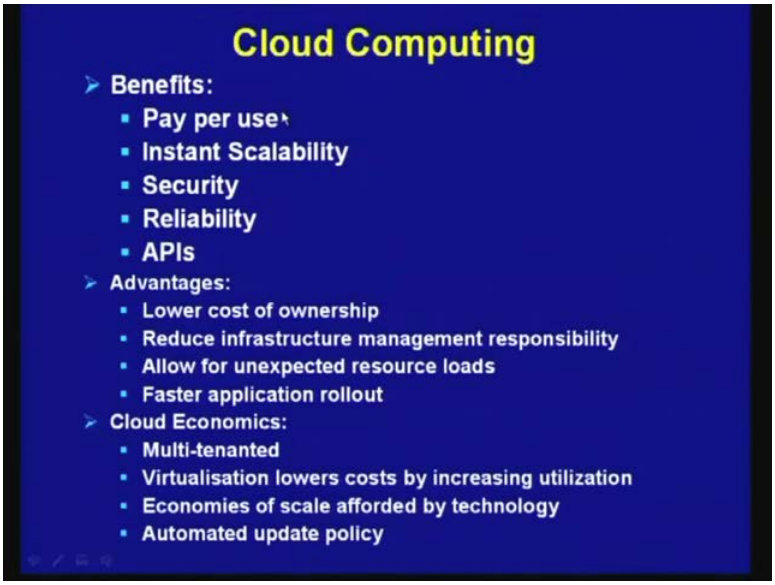
(Refer Slide Time: 49:21)



And there are three important segments of applications, first one is the known as applications, so applications software has a service, so which is known as SaaS in short. SaaS provides a complete turnkey application such as enterprise resource management throughout through the internet, so what you are doing here, you are using an application, complete application through internet provided by a third party. Second type of second segment of cloud computing is platform, so platform has a service, so which is known as PaaS.

So, PaaS offers full or partial application development that users can access, so it a platform is provided to you, for full or partial application, and that can be used and with the help of this segment, second segment platform segment. Third is your infrastructure, so infrastructure has a service or IaaS, so a consumer can get service from a full computer infrastructure through the internet. So, without procuring it, you are able to get full service, full computer infrastructure through the internet provided by a third party. So, these are the three main segments, which are mainly available through cloud computing.

(Refer Slide Time: 51:13)



Now, let us look at the different benefits and advantages of cloud computing, so the benefits are you pay per use that means, as much as you use you pay for only that. And you can have instant scalability, as soon as you are requirement increases, you can scale it very quickly. And you can scaled on very quickly in an instant, because you are not deploying it, deployment it is deployed by others, so based on your requirement you can scale it up and scale it down, and only you pay for your use.

Then you can have high security, high security is provided through cloud computing, and you can have high reliability as well, because when it is in cloud computing as you have seen, you have got large number of computers and systems, and it gives you high reliability. And it also gives you high API's, so these are the benefits of cloud computing, and the advantages of cloud computing is listed here, with provides you

lower cost of ownership. Since, you are not deploying anything you are not buying any infrastructure hardware or software, your cost of ownership is very small.

So, it provides you lower cost of ownership and reduce infrastructure management responsibility, so infrastructure management responsibility is also not with you, since it is not part of you, your not really performing the infrastructure management. Infrastructure management is done by third party your simply using it, and this is one advantage and it allow for unexpected resource loads. So, whenever you have got unexpected resource load, you have got large number load, you can very easily scale it up, to meet the a load and faster application roll out.

So, because of this advantages, you are working on an application by using the infrastructure and resources of cloud computing, using cloud computing you can complete it and you can roll out you application very quickly. So, these are the main advantages and this cloud economics is best on multi-tenanted, so that means, you can have multiple resources, your using as tenants and virtualization lower cost of cost by increasing utilizations.

As I mentioned hardware and software or virtualized, and bit lowest cost by increasing utilization of the resources, and economic subscale afforded by technology. So, since large number people are using it that the economy as a result cost is divided among a large number of users, so economic scale afforded by technology. So, technology itself is providing this sharing of cost by a large number of users, and automated update policy.

So, that means that as, and when as the requirement keeps on increasing, the infrastructure is also upgraded to satisfy the requirements of users, whether it is hardware resource or software resource, the resources kept on increasing. So, these are the different aspects of cloud computing, so to summarize in this lecture we have discuss three important types of computing. First one is cluster computing that is primarily used for high performance computing.

And we have seen computers are within a small geographic area that is LAN, on the other hand grid computing and cloud, in case of grid computing and cloud computing resources are distributed through internet. And of course, in grid computing it is done in one way and in cloud computing it is done in one way, and because of many benefits and advantages of cloud computing, it is becoming increasingly popular.

Thank you.