Cloud Computing and Distributed Systems
Dr. Rajiv Misra
Department of Computer Science and Engineering
Indian Institute Of Technology, Patna

Lecture - 01
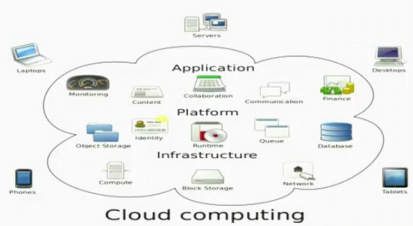**Introduction to Cloud Computing**

Introduction to Cloud Computing. Preface, content of this lecture you we will discuss a brief introduction to cloud computing.

(Refer Slide Time: 00:20)

And also focus on the aspects such as; why clouds, what is a cloud, what is new in today's cloud and also distinguish cloud computing from the previous generations of computing system that is, the distributed system.
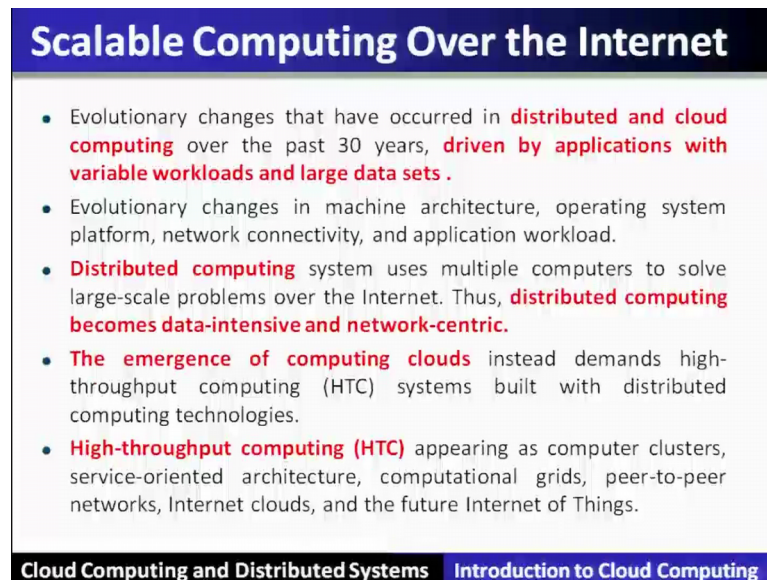
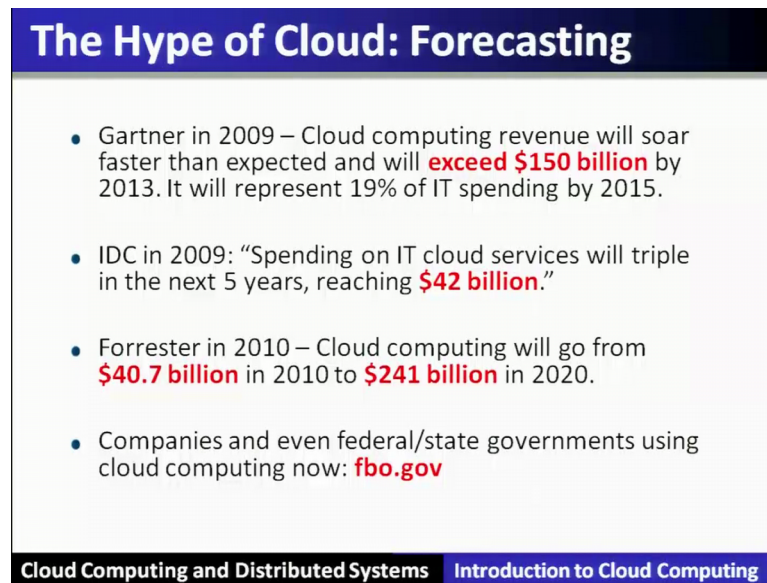(Refer Slide Time: 00:40)

Scalable computing over the internet, so the evolutionary changes that have occurred in distributed and cloud computing over past 30 years are driven by the applications, with variable workload and large data sets. These evolutionary changes are happening in the machine architectures, operating system, platform network connectivity and application workloads. The distributed computing systems uses multiple computer to solve large scale problems over the internet. Thus, distributed computing becomes the data intensive and network centric.

The emergence of computing clouds, so instead the demands of high throughput computing systems build with distributed computing systems has led to the emergence of the need of the cloud. So, in high throughput computing we will see that the systems which appears they are basically the computing clusters, service oriented architecture, computational grids, peer to peer networks, internet cloud and the future of internet of things.

(Refer Slide Time: 02:12)

Let us see the hype of cloud which was forecasted around 10 years back, and using that particular predictions we will see how far these predictions are towards moving the cloud further. Gartner in 2009 has predicted that the cloud computing will soar its revenue much faster; than the expected at that point of time and we will exceed cross a huge revenue mark. And it will represent the 19 or 20 percent growth of IT's spending in terms of the cloud computing by 2015.

Similarly, IDC in 2009 also predicted the same thing that; spending in IT cloud services will triple in the next 5 years. And Forrester in 2010 has predicted also the same thing; that is the cloud computing will go as far as spending is concerned in 2010 to 2020; crossing a several that is 5 folds' increase. So, companies and even new federal governments are also using the cloud computing.

(Refer Slide Time: 03:48)

So, with this particular forecast let us see; what is the current status, how many key players in the cloud is available as the providers. The first one is called as Amazon Web Service, which is a most prominent cloud provider. So, it provides the cloud services in 3 different types; the first one is called Elastic Compute Cloud, the second one is S 3 that is Simple Storage Service, the third one is called EBS Elastic Block Storage. The second cloud provider which is also well known is called Microsoft Azure, and Amazon and Microsoft Azures normally they provide similar kind of cloud services.

The third one is called the Google Compute Engine that is also known as App Engine, that is also a cloud provider. Besides these 3 prominent cloud providers there are several other cloud providers such as; Rightscale, Salesforce, EMC, Gigaspaces, 10gen, Datastax, Oracle, VMWare, Yahoo, Cloudera and there are 100 many more in this particular arena.

So, there are categories of clouds which are available, they are called as public cloud and the private cloud. Public clouds are accessible internal to the companies and they are also managed internally by the company, and all it is basically like energy, usage. It is maintenance all are owned by the company itself; hence it is called a private cloud. It is not available it is not accessible by outside people.
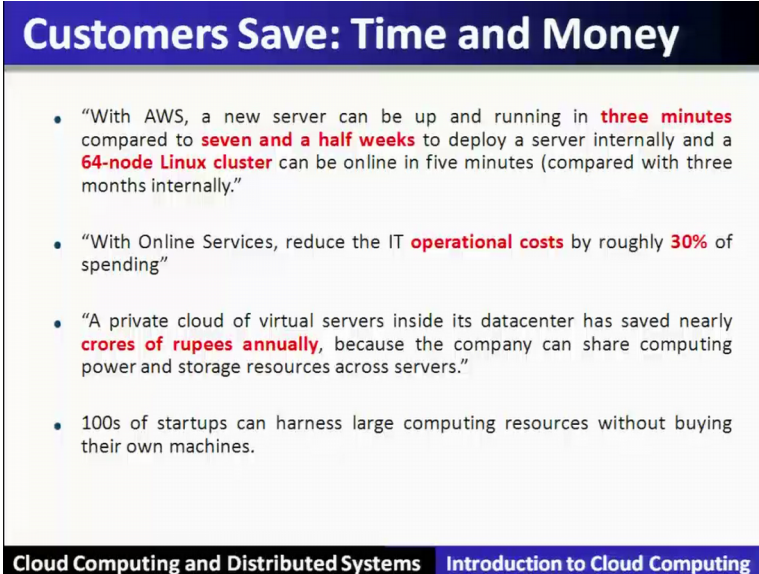
The other type of cloud is called a public cloud, which provide the services to any paying customer; that means, it is open for anyone who want to use the cloud by paying the cost, that is why it is called a public cloud. The cloud provider has to basically maintained, then bare the cost of energy and so on and so forth. Only the customers who want to use it they have to just pay as they use it. The example of public clouds are the Amazon S 3, Amazon S 3 is a Simple Storage Service will store arbitrary datasets, and the users has to pay the money as per the amount of space which are basically rented are used for the storage, that is; in terms of GB per month.

The second kind of public cloud is given by the Amazon as EC 2 that is; Elastic Compute cloud, and this particular elastic compute cloud will provide the compute services to the client. So, as per as any user can upload and run an arbitrary operating system images. And based on this several operating system, which are basically given and different applications can basically be used be run on this kind of system.

These particular different operating systems, images and applications which runs will require now the CPU. So, therefore, the user has to pay, how many CPUs they require; that means, instead of numbers, but they have to pay as per the CPU hour, which is being used by the applications.

Similarly, the third kind of service which basically is categorized under the public cloud is, example is called Google app engines; here the users can develop their applications within this app engine framework and they upload their data that will be imported into their format, and it can run. So, it will give you that Google app engine gives more flexibility in terms of the directly the programming customer can do. And they can use the entire framework to solve their applications, and they can pay the money accordingly the use.

(Refer Slide Time: 08:20)

As far as the customers are concerned, in this cloud scenario the customers will save the time and money how that we are going to explain over here. So, if AWS is being used, then basically a new server can be up and running and within a fraction of minutes that is; 3 minutes compared to several weeks and months to purchase the server, and then basically put it into the service. All these particular cost of invoicing purchasing and installation will be now reduced and only 2 or 3 minutes are required to install a new server and run. So, time is saved if the cloud is opted as the method of computing instead of owning the own servers.

Another example is regarding with the online services will reduce this operational cost by around 30 percent of the spending of the internal company, why because operational cost is not at all required. The only amount of money it has to be paid whatever as per the use. Therefore, it is a saving of money also if the cloud is used for computing purposes. A private cloud of the virtual servers inside the data centre has saved nearly crores of rupees annually because company can share; computing power and storage resources across the servers.

So, again this is also going to be very cost effective usage, if even if the private cloud is maintained inside the data centre. We will see the economics when it is required to be go for the private cloud versus the public cloud. Also there are various startup companies they can harness large computing resources without buying their own machines. This also is based on the economics calculations, whether to go for a your own private cloud or a own or basically use the public cloud, but as far as we will see all these things why, so most many options are available; and these options open up new arena of a cloud computing. So, what is a cloud?

(Refer Slide Time: 10:59)

What is a Cloud?
- Advances in virtualization make it possible to see the growth of Internet clouds **as a new computing paradigm**.
- i.e. dramatic differences between developing software for millions to use **as a service** versus distributing software to run on their PCs."

History:
- **In 1984, John Gage** Sun Microsystems gave the slogan, "**The network is the computer.**"
- **In 2008, David Patterson** UC Berkeley said, "**The data center is the computer.**"
- Recently, **Rajkumar Buyya** of Melbourne University simply said: "**The cloud is the computer.**"
- Some people view **clouds as grids** or **clusters** with changes through virtualization, since clouds are anticipated to process huge data sets generated by the traditional Internet, social networks, and the future IoT.

**Cloud Computing and Distributed Systems** | **Introduction to Cloud Computing**

Now, here we will see that the advances in the virtualization makes it possible, the growth of internet clouds as a new computing paradigm that is; there is a dramatic difference between developing a software for millions to use as a service versus developing a software and distributing it to run on their own PCs. So, the architecture in

a cloud computing is now slightly changed, where the software will be given as a service to the millions rather than the software is to be distributed to run on their PCs.

So, the cloud has changed the new paradigm and let us trace back through the history. In 1984 John Gage of Sun Microsystems gave the slogan that the network is the computer. Similarly, later on in 2008 David Patterson of UC Berkeley has said that; the data center is the computer recently Rajkumar Buyya of Melbourne university simply said the cloud is the computer. So, just see the way the paradigm is shifting the definition of computer is not changing from network to the data center, now it is the cloud.

So, some people view the clouds as a grid or a clusters, with changes through the virtualization these clouds are anticipated to process huge data sets which are generated by the traditional internet, social networks and future IOTs.

(Refer Slide Time: 12:46)
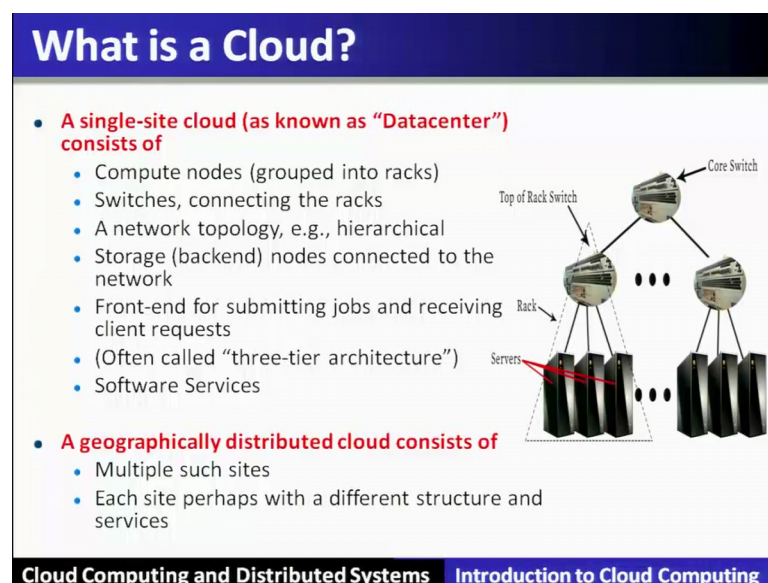
So, we will see the inside what is there in the cloud. So, if you go inside we will see; that there are two kinds of setup you will see inside the cloud. The first one is called single side cloud that is within one premises. That cloud is called the datacenter, which comprises of the compute nodes which are grouped into the racks, which are shown here these are the servers, which are grouped into these particular racks, this compute nodes are there sometimes they are also called as server.

Then comes the switches which are connecting these racks. So, every rack will have a top of the rack switch, which is basically mentioned over here and they are connecting all the racks. Then comes this particular network topology which will be of 2 level, and within this particular rack will also be the storage backend nodes connected to the network so; that means, there are nodes also within that particular rack which are primarily meant for the storage purposes. They are having the SSTS within it, so basically used primary for the storage.
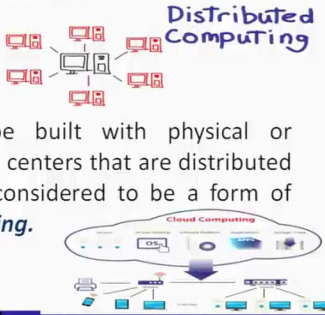
Now, frontend will be there for submitting the jobs and receiving the client requests. So, often this is can be treated as 3 tier architecture. Here there is a core switch which we will connect all the different top of the rack switches. So, this particular hierarchy and there is a software services, which will basically will be used to run the applications on this kind of structure, which is basically nothing but a datacenter with the clusters within it.

Now, there may be there are different cloud providers, which has deployed this kind of set up at more than one sites they are called geographical distributed clouds. And which comprises of multiple search sites that is; multiple search datacenters which are connected together and each site perhaps with a different structure and the services running within it they can communicate over the fast network.

(Refer Slide Time: 15:09)

So, that was some of the interval the description of the cloud, what comprises of the cloud. Now, this cloud what is basically the computing paradigm which makes the distinction as a cloud computing that we will see. So, there is a wide overlap between the cloud and the distributed computing. Distributed computing comprises of multiple autonomous computers having their own memory and they communicate through the messages called message passing. As far as the cloud is concerned cloud can be built with the physical or virtualized resources over the large data centers that are distributed systems.

So, basically on these particular distributed systems, the virtualization of the resource will create pool of virtualized resources. And this pool of virtualized resources can be allocated to the applications. And therefore, the cloud computing is having an overlap with a distributed system and also a flexibility or elasticity, which is called in terms of computing resources; so, the cloud computing is also considered to be a form of utility computing or the service computing.

(Refer Slide Time: 16:27)
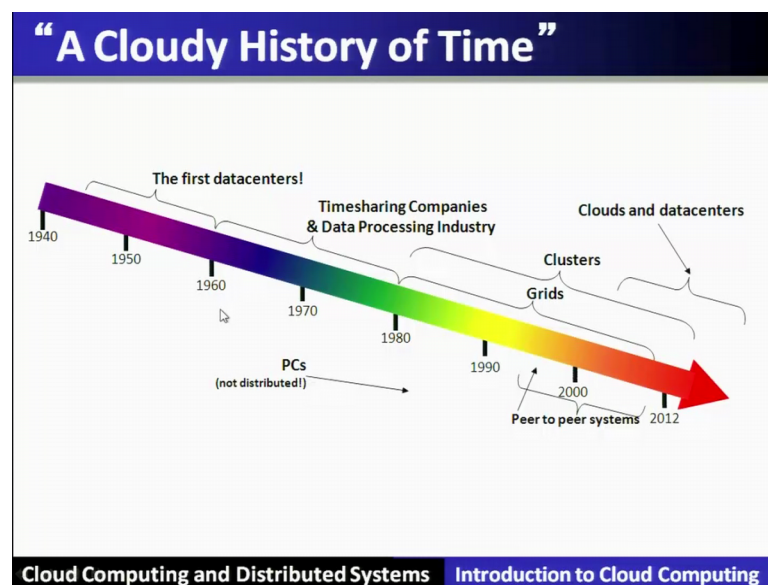
Let us trace back the history of the development of a cloud system. So, 1940s we will see we have seen that; here we have seen that the ENIAC the bigger the big data centers like ENIAC and ENIAC system was installed. And these particular systems in 1940s were housed in a big room. And that was the data centers we call those big rooms, which are full of CPUs called data centers, but primarily with a slower, slower computing facilities

which we have right now. Then afterwards then came the timesharing companies and the data processing industries were transformed into a timesharing system where terminals and the PCs terminals were used to express those systems.

And also the data if it is quite large then it is given in the form of a punch cards. And that industry was called data processing industry. Then there is a slight change after 1980s and the systems were now become the PCs personal computers and personal computers were given to the people directly to use it. And at the same point of time the grids were also evolved clusters were also evolved. Then using these particular systems, the peer to peer systems were formed they are the precursor or precursor to the current cloud computing systems.

Now, the current cloud computing systems are basically having the same setup, that is called datacenters which were there in 1940s so and 1960s, so the same cycle is being repeated, but with the different notions of the computing, so that more data or a big data can be computed into this kind of datacenters.

(Refer Slide Time: 18:37)
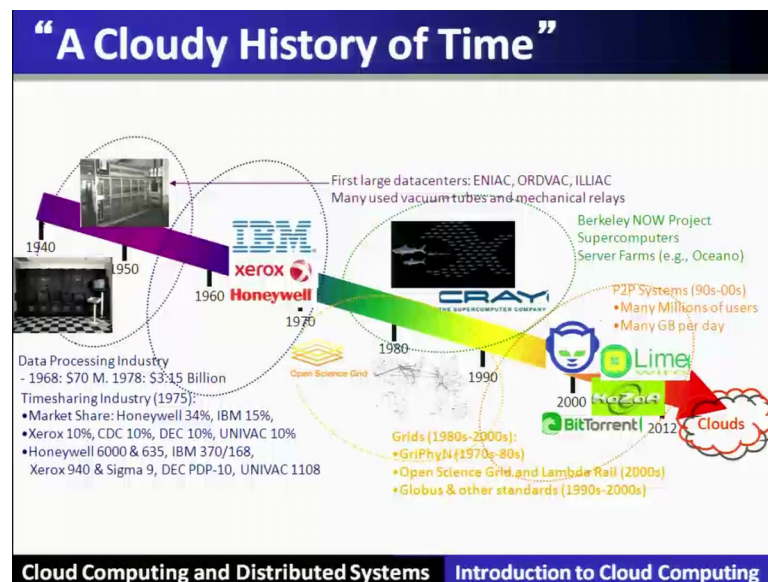
So, this is what is being summarized over here that; the precursor to the clouds are basically the peer to peer why because many PCs were available and they were, they were connected together to form a computing, and a cloud is basically further advancement of these kind of systems.

Now, as far as the amount of data and the flexibility in the applications for these resources, let us to the scalable computing trends and the technology is also evolving around that. So, we will see how the trend is in the technology perspective has taken up the shape and has given, now the birth for the cloud computing. So, as far as the hardware is concerned in the hardware we have seen the scalability or the growth that; it was and we such that the storage after every 12 months it was envisaged that it will be doubled.

Similarly, the bandwidth after every 9 month will be doubled similarly the compute CPU compute capacity also every 18 month will be doubled. Doubled in the sense with the same cost you will get the doubled the capacities, double the speed and double the bandwidth; that means, the doubling phenomena. So, what is the law behind this doubling of the periods? So, Moore's law indicates that the processor speeds doubles every 18 months. Although there is no basically doubling in the terms of the speed, but horizontally the development is taking place. So, number of course, are being packed more in a corresponding chip that is the trend now.

Similarly, the gilders law indicates that the network bandwidth has doubled each year in the past. Similarly, we see that earlier the bandwidth was in kbs kilobits per second. In 2015 we will see a terabytes per second, was a link speed of the same amount or a same cost. So, there is a tremendous increase and it is being followed the principles of a

doubling period. Similarly, the disk capacity today's PCs have terabytes for more than 1990s super computers.

(Refer Slide Time: 21:13)

So, this all is moving towards the reality of utility computing. Let us see what do we mean by duality computing. So, aiming towards autonomic operations that can be self-organized to support dynamic discovery major computing paradigms are composable with quality of service and service level agreement SLAs. In 1965 MIT's Fernando Corbato of Multics operating system envisaged that; computing facility like operating systems like a power company are basically work like or like a water company.

So, power company or a water company works like the plug and play; that means, the power is available in the homes in the form of a socket, whenever required you can use the plug and continue to use it without having knowing the problems or the production at the power stations. Similarly, the water companies problems without knowing it, you can open the tap water will come out, using the same concept of utility computing.

So, computing also should be provided in the similar manner if it is then it is called the utility computing. That is the thin client can plug-in into the computing utility and play; that means, one can use the compute and also can run the applications. So, cloud in some form is realizing the hope of the computing utility computing. So, utility computing focuses on business model in which customers receive computing services from a paid

service provider. So, here all the grid oblique the cloud platforms are regarded as the utility service providers.

(Refer Slide Time: 23:14)

Features of today's clouds, so there are 4 different features which will categorize the cloud or the applications which are basically the cloud problems. The first one is massive scale, very large data centers contains tens of 100s of 1000s of servers. And you can run your applications across as many servers as you want and as many servers as your application will scale, that is called massive scale. We will see in more details of what do you mean by this massive scale and in terms of the cloud problem.

The second aspect or the feature which will classify as a cloud problem or a cloud computing is called on demand access. So, on demand access means pay as you go pay as you use; that means, it is different from upfront cost, upfront cost means you have to pay in advance and then whether you use that up to that level or not, that is not that is in contrast to that. So, on demand access the pay as you go, so these particular model also classify the problem into the cloud problem.

Third feature which classifies the problem into a cloud problem is called Data intensive nature. So, what was megabytes earlier now has become a bigger size that is in a terabyte Petabytes and Zettabytes. So, this size of data is growing and if the data size is quite large, then those problems falls into the cloud problems. Examples are daily logs forensics reports then weblogs, which will continuously generate the data that becomes

of that size data and it has to be solved in a data intensive nature, that is the computing of that category of that system is required that is called a cloud computing.

4th feature is called a new cloud programming paradigm. So, the problems of the big data or a large scale data or a data intensive nature of applications, require a new cloud programming paradigms for example, map reduce and it is open source version is called a new programming paradigm, which is used to solve this particular problems. So, this new programming paradigms also are classified as one of the features of the cloud problem.

Another thing is called the key value store, if it is then the systems like Cassandra is being used. Similarly, if the database is in the form of NoSQL then MongoDB is basically the programming paradigm which is being used. So, newer programming paradigms are available and if they are required then basically the problem is categorized in this particular today's cloud. Now, if we will see that if one or more of these above features are available then only we can classify the problem into a cloud computing problem.

(Refer Slide Time: 26:52)

Let us see in more detail what do you mean by the first feature of a cloud computing that is called a massive scale. Take for example, the Facebook application, Facebook application as of 2012 we have seen this particular data that there were 30000 servers were deployed in 2009, which has grown up to 60000 servers in 2010. That is in one year

it was doubled number of servers and in 2012 it is 118000 servers are deployed. So, the scale is basically keeps on changing and it has become 180000 servers are used to run one application that is Facebook is a massive scale.

Similarly, the Microsoft, Microsoft in 2008 were using 150000 machines and that growth rate was 10000 machines per month. And we can see that 80000 different servers were running one application which is called Bing. Similarly, in 2013 Microsoft cosmos application required 110 thousand machines. And those machines that many number of machines were deployed in 4 different regions. So, that is basically called as the massive scale and this massive scale is required to serve these applications, which is termed as the cloud. Similarly, Yahoo in 2009 has 100000 servers and that splits up into the cluster of 4000.

So, that is there are different sites of at most 4000 servers and they were together if we see that they becomes 100000 servers which runs this Yahoo service. So, Yahoo is basically providing the Google or using the cloud service. The next one is Amazon EC 2, we see that in 2 2009 40000 machines were required to run this particular system or application EC 2. And each machine was basically of 8 core systems. Similarly, eBay required a 50000 machines to run the applications, HP 380000.

So, as far as the Google is concerned it requires lot many number of servers the total numbers are not disclosed by the Google, but it is basically known that it is quite large than any of the above companies which we have. So, it is basically a massive scale. So, this is the first requirement of the cloud problem.

(Refer Slide Time: 29:49)
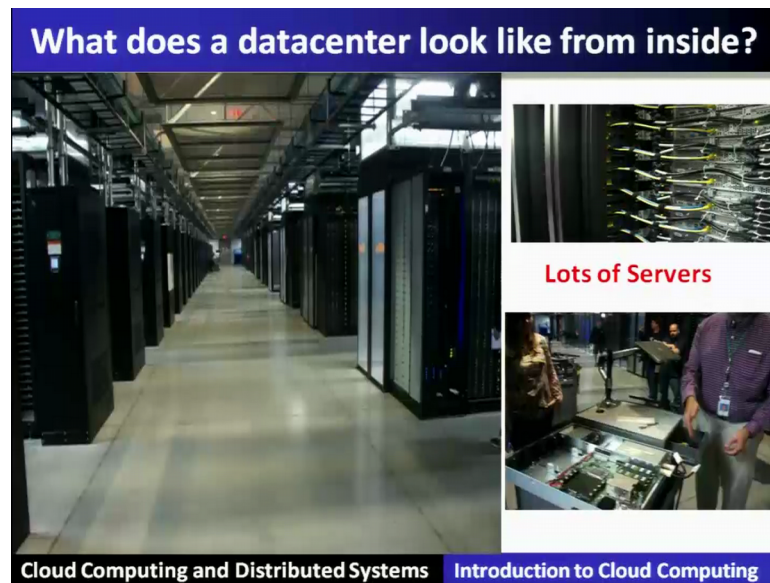
Now, what is there inside the massive scale that we can see what is there inside the datacenter. So, at one side and at several sides these datacenters will house lot of servers racks and they are all connected together. So, that you can see in this particular room this is called a data centre it has lot of racks full of racks the entire room is filled, and all these racks are connected on the right side you can see this back side of the rack, they are all servers which are interconnected with each other.

And, within inside every server you can see the boards, they are servers are nothing but they are basically in the form of a blades or in the boards, all are fitted within this particular rack and they have being powered there have being communicate they are been connected through the network

(Refer Slide Time: 30:46)

Such a datacenter requires huge power to basically run and, and also this power is being generated through the power stations and also when so much of power is required within one room, so lot of heat is being generated. So, how to cool it so all these are basically the requirement for the maintenance and it requires the cost, how to reduce this particular energy is one of the challenge here in the cloud computing datacenters. So, the water which is used to cool down, and so the annual water usage that is water usage is also measured as annual water usage divided by the IT equipment energy which is being used if this particular parameter is low then it is good.

Similarly, the power utility is also measured by the total facility power divided by IT equipment power; if it is low then it is good. So, Google has shown that his particular data center is achieving 1.11 less than 1 is not possible, but it is very close to 1 so; obviously, it is trying to use as much as it is power drawn for a computing purposes without any much energy wastage.

(Refer Slide Time: 32:11)

So, the cooling there are different types of methods are used to cool this datacenter and some of them are like air sucked in and water is also combined with purified water. And water moves the cool air through the system. There are various methods of cooling which are used in the datacenter of a cloud.

(Refer Slide Time: 32:32)

Now, the second feature of a cloud is on demand access. So, in the industry terms it is being classified as AAS classification so; that means, hardware as a service IaaS means infrastructure as a service, then PaaS means platform as a service, SaaS means software

as a service. So, on demand axis is one of the important features of a cloud problem, on demand means; that you are not buying you are renting or you are paying as you use. So, for example, AWS elastic compute cloud EC 2; that means, you can pay as you use the CPUs for it. How many CPUs are required, how many CPU per hour is required for the application only that is being paid off by the customer.

Similarly, AWS also provides another cloud service that is called storage service S 3, AWS simple storage service. So, it has to be paid as per the use; that means, GB per month. How much space or a storage space per month you use and that amount is to be paid that is called on demand access to the resources like computing storage and therefore, it is been classified. The first one is called hardware as a service; that means, you can get access to the bare bone hardware machines and do whatever you want with them; that means, your own cluster can be owned by someone else as a service.

But it is not a good idea because of the security risks. Therefore, the another one which is called infrastructure as a service is more popular than hardware as a service, and is being provided for the public to be and many companies are use are family giving infrastructure as a service to the customers. So, that access to the flexible computing and storage infrastructure in infrastructure as a service and this is done through the virtualization. So, virtualization has achieved this infrastructural service, and infrastructure as a service has subsumed the hardware as a service. So, hardware as a service is not being used up in the industry, but infrastructure as a service is used and that also covers hardware as a service within it.

For example, Amazon web services AWS, EC 2 and S 3 is example of a infrastructure as a service, OpenStack is also an example of a infrastructure service. Eucalyptus is also an example RightScale, Microsoft Azure and Google cloud they are example of infrastructure as a service.

Now, another kind of AAS classification for on demand axis is called platform as a service. Now, get access to the flexible computing and restorage infrastructure coupled with the programming platform. So, often they are tightly coupled so; that means, the programming paradigm is given and the people the users can use the programming to basically run their applications examples are Google app engine.

This is not as flexible as infrastructure as a service, but it is easy to use platform as a service. Another way of AAS classification is called software as a service. So, get access to the software services when you when you need them. And this will subsume the service oriented architectures. So, given that software as a service is available. So, the service oriented architecture is being subsumed the example of a software as a service are the Google's doc and MS office on demand these are the software as a service.

The third important feature for a cloud problem or a cloud computing classification is called data intensive computing. This is in contrast to a computation intensive computing earlier there was a computation intensive computing. So, the data was a small and which has to be computed very fast.

So, therefore, MPI based high performance computing cluster or grids were formed and also the development of supercomputer was to compute the data very fast. So, that is called computation intensive computing, but the trend is now changing, here the applications have a large data and data cannot be moved, where the compute nodes where the computations are, but the computation has to move where ever the data is required for it is computation, that is why the size of the data is too big. And that is why it is called a data intensive computing.

So, with the typical data intensive computing that is one of the key features of a cloud computing systems requires to store the huge amount of data in first hand at the datacenter. The second is use the compute node nearby; that means, the data cannot move because it is the big size, the compute nodes nearby can computed. So, compute nodes runs the computation service that is called data intensive computing.

So, in data intensive computing the focus shifts from the computation intensive computing to the data intensive computing. So, that is the focus shifted from computation to the data. Here the CPU utilization is no longer the important resource

metric, but instead the IO input output that is the disk and also the network is important why because; in a data intensive computing IO is very important similarly the network is important.
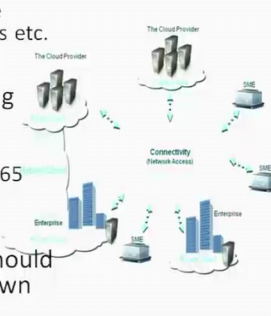
(Refer Slide Time: 38:54)

Now, the next important features of the cloud problem is called new cloud programming paradigm. It has to provide the easy to write and run highly parallel programs in the new cloud programming paradigms. For example, the Google provides a new programming paradigm which is called MapReduce and Sawzall. So, Amazon also provides elastic MapReduce service, where you have to pay as you go. Google also provides the MapReduce in the same manner. So, map Google also uses the MapReduce in the form of indexing, for indexing it requires a chain of MapReduce of 24 MapReduce jobs.

So, just see that the MapReduce solves a big problem or the data intensive computing. Similarly, the Yahoo also has used the Hadoop that is a open source version of the MapReduce and it is own version that is called pig. Facebook also uses the hadoop and the hive where in 300 terabytes of total data size is being basically computed or being processed. Another thing is called a new programming paradigm is NoSQL, which is in contrast to the MySQL which is of industry standard. Similarly, the key value store that is called Cassandra is also 2400 times faster than MySQL.

(Refer Slide Time: 40:35)

So, we can see that as far as the cloud is concerned there are two types or two categories of cloud public and private. So, as you know that the private clouds are accessible only within the company and public clouds basically are being provided as a service to the customer.

(Refer Slide Time: 40:54)

Now, the question about whether to use the private cloud or to use a public cloud is basically a matter of economics. So, for example, if let us say we have taken example of a medium size organization, which runs it is computing services for let us say M months.

So, the services in requires 128 different servers and 524 terabyte of a space. Now if it is outsources using Amazon AWS services on a monthly basis. So, let us compute the cost. So, for a storage S 3 will cost let us say about 62 K dollars. And CPU will be costing around how much that is comes out to be 1024 times 24 times 30 that comes total comes out to be 136 K dollars for the outsourcing of the entire computing infrastructure.

Now, in contrast to that if you want to own a private cloud and let us understand the cost of that storage. If you purchase 349 K dollars divided by the total number of months you are going to use it. Similarly, you have to add some more cost for it is maintenance in the terms of the man power system administration and so on. If you compute what you will see here is that this particular outsource is if it is equated to this particular purchasing it over a particular period of months you will obtain a breakeven analysis.

And in the breakeven analysis if you see in the slides that if the number of months is more than 6 months, then it is better to own the infrastructure that is go for a private cloud. And if the number of months is more than 12 months for the overall then only you can go for the private cloud. If it is less than 6 months and less than 12 months for the overall, then the breakeven says that you should not go for the private one you have to go for the public one. Therefore, the startup companies which uses such a infrastructure, but for a little period of time maybe less than one year for the experimentation all it is better for them to go for the cloud, that is why clouds are becoming more popular for these companies.

(Refer Slide Time: 43:49)

Conclusion, so clouds build on many previous generations of distributed systems. So obviously, you have to see that the development of distributed systems with the virtualization has built the new generation of the cloud computing systems. These cloud computing systems have the features or the characteristics in the cloud problems, they are called massive scale, on demand access, data intensive computation at a programming paradigm.

Thank you.