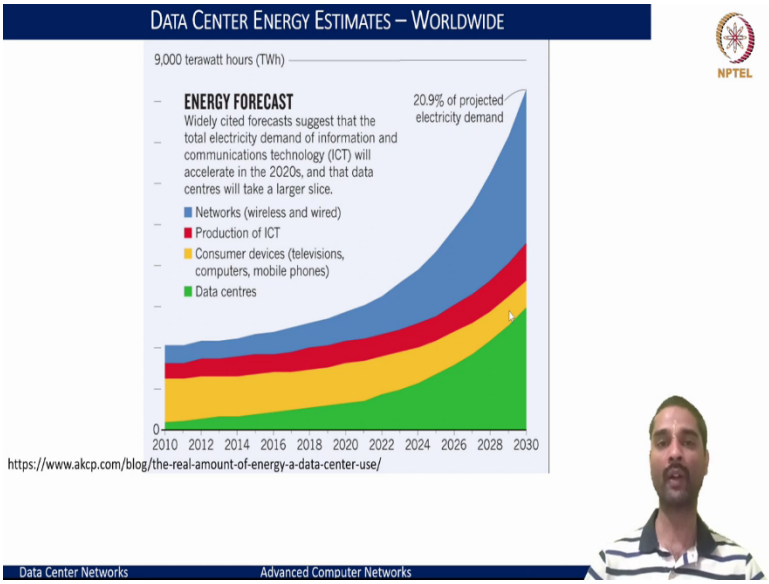
Advanced Computer Networks
Professor Doctor Sameer Kulkarni
Department of Computer Science Engineering
Indian Institute of Technology, Gandhinagar
Lecture 60
Green and Sustainable Data Centres

(Refer Slide Time: 0:17)

So far, we looked at the data centers in terms of trying to understand the components, the way the data centers are classified, the aspects in which the network topologies are built, and the key protocol and architecture innovations that happened in the space of data centers. And we said the amount of data centers across the world has been increasing sharply. So, we also need to look at the energy prospects of what the data centers demand and where it would lead us in the future. And in essence, we need to understand the greenness quotient and sustainability quotient of data centers.

(Refer Slide Time: 0:49)



So, there was a survey that was conducted, and it showed like how there has been a variation in terms of the energy estimates of the world in terms of how or which sectors are they being driving and what has been the contributions of data centers in it. And as of 2022, if we look around the world, we have more than around 8000 data centers, and most of them almost 30 percent of them are in U.S, and in India, we have almost around 150 of the data centers that are co-located data centers that have been set up.

And each of these data centers, if we club together and say the amount of servers that they host and across the world if we combine all of them, we expect somewhere around 18 to 20 million servers around the world in data centers as of the date and this numbers might have been grown higher because it was like almost 2 years old data and each of the servers we said they have a power rating that means there is a lot of power that you need to drive these servers, drive these racks and travel, drive the entire data center.

And in 2020, data center industry consumed around almost 200 to 400 terawatt hours of energy and, that is, basically equivalent to almost 2% of the worldwide annual energy consumption was taken up by the data centers. And similar analyses were done across the globe some in India, some in the US all shared similar prospects in terms of how the data center industry has been becoming more and more energy-hungry.

And from India, what I could see there was an IBM Global report which said almost 40% of the emissions carbon emissions that happened in ICT were due to data centers, and roughly 50% of

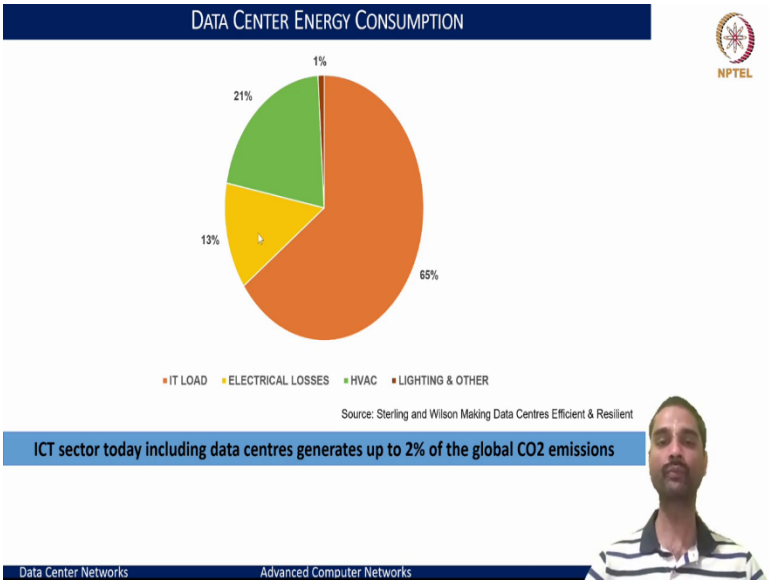
those emissions when we contribute towards the ICT were due to the data centre's power losses or system losses and cooling loads that meant, deficiency in terms of how you drive these powers affects a lot.

And recently, as per MIT, the Ministry of Information Technology, even the size of the digital economy is growing from almost 200 billion around roughly 2018 to almost 1 trillion, which we are expected to reach around 2025 means that we have a lot more necessities to bring in and the data center landscape in India is only going to explode.

And given this explosion that we may see in the data centers' work, we need to understand how we are able to adapt and set the data centers' energy drives. So, the start over here from the AKPC tries to say how it forecasts the variation in terms of the overall energy that we see in the world to the consumption of the energy by different key sectors, and we can see as of 2010 to 2015, most majority of it was by the consumer's end devices, and then the things started to grow over the data centers.

And we can see as we reach towards 2024, 2025, and almost around 2030, more than around 30 to 40% will be contributed towards driving these data centers. And this is a significant share if we have to think of the 9000 terabytes hours we would be driving data centers at around 3000 terabytes hours needed across the globe. Hence, we need to understand what are the key components within the data center that require this energy or power and how, if at all we can make these things to be more sustainable and drive them towards a Greener technology.

(Refer Slide Time: 04:32)



And there were a lot of studies that were conducted, and we see diverse opinions in terms of how the overall energy consumption within a data center varies, but the key contributor is primarily the IT load, that is, basically, the IT infrastructure, the servers, the switches the middleboxes that have been hosted within a data center and the power to drive them.

And in this plot here, which was shared by Sterling and Wilson on the data centers energy and efficiency and resiliency report in 2019 entered about almost 65% going towards the IT load, and 21%, the other major share going towards just the cooling part of the HVAC, what we studied is the heating ventilation and air conditioning that we would want for the data center cooling center, the CRAC.

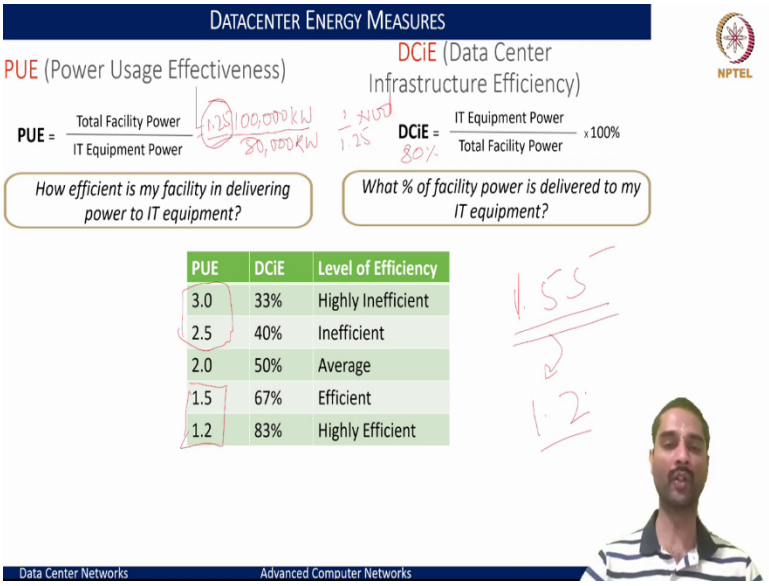
And 13%, which is still a major share, was used up for electrical losses, that means, we need better technologies to make up for these losses, better technologies to make sure that we are able to do better cooling or efficient cooling and also the IT infrastructure itself if can be made to lose much lower power, and this is where the things would shape for the data center energy.

Nonetheless, these numbers are a bit debatable because there were academic studies that showed the cooling itself to contribute to roughly 40 to 45% and IT workload to constitute around 50%, and the rest in terms of networking, in terms of electrical losses etc contributing to the remaining 10%.

And these numbers are just the ball marks that we can think of and see what mechanisms we can try to build around all of these aspects, especially the IT load, that is, the servers and the

infrastructure things, including the lightning, etc that will cost you towards the data center energy and the cooling system that we would want to deploy as well as mitigating the electrical losses. So, these would be the key aspects to look for if we have to think of making these data centers more sustainable and green-oriented.

(Refer Slide Time: 06:34)



In this aspect, it is all very essential that we now try to dictate and define the key metrics through which we would be able to define the effectiveness of the data center energy consumption, have a measure or leftover, how energy is being conceived by the data centers, and what would make it better.

In this aspect, in the early 2000s, there were a lot of benchmarking parameters that were being set widely in Green Community the data for the data centers, and the most effective one that we use today is what we call as the Power Usage Effectiveness or the PUE and PUE and its reciprocal also will touch upon its the DCIE or the Data Center Infrastructure Efficiency are the most widely adopted benchmarking standards.

And these were, like I said, set up by the green grid to help the IT professionals determine how the Energy Efficiency of a data center can be measured and monitored so that things can be improved and the PUE is basically the measure of energy efficiency of a Data center and what does that mean the IT equipment part of the power that is essential is the core component which is taken out from a subset of IT power. That means when we feed the total power, it is going to

be channeled some for IT equipment and some for other aspects. So, in a nutshell, PUE measures the efficiency of the data center facility towards delivering the power to the IT equipment.

And if we think of a reciprocal of it that is the one over PUE, it translates to saying what percentage of the total data central power is being delivered to the IT equipment expressed as a percentage, and that is the other key metric that is being used that is the DCIE and this is exactly 1 over PUE multiplied by 100. And the two metrics try to give a guideline in terms of how the power is being utilized for the data center.

So, if we consider an example where we say that a data center facility uses roughly around, let us say, 100,000 kilowatts as a total facility power, and of this, if we say that almost around 80, 000 of the kilowatts is being used to track the IT equipment, then the PUE will in essence mean 100,000 kilowatts/80,000 kilowatts roughly translating to 1.25.

So, you can say the PUE index for such a data center would be around 1.25 here. And likewise, if we translate that to DCIE, that would be (1/1.25) x 100, and this would translate to basically saying that it is 80% as the effectiveness in terms of how much of the total power has been delivered towards IT equipment. And these numbers vary based on the kind of data center in terms of what all that is facilitating and how much of the power is being used by the IT equipment.

So, the metrics were defined in terms of if the range of a PUE falls in a particular category, and likewise, the DCIE would vary, and we can categorize those as efficient and inefficient, and a PUE of a number 3 or greater than 3 would be categorized as highly inefficient data centers that need to work upon and improve the power ratings.

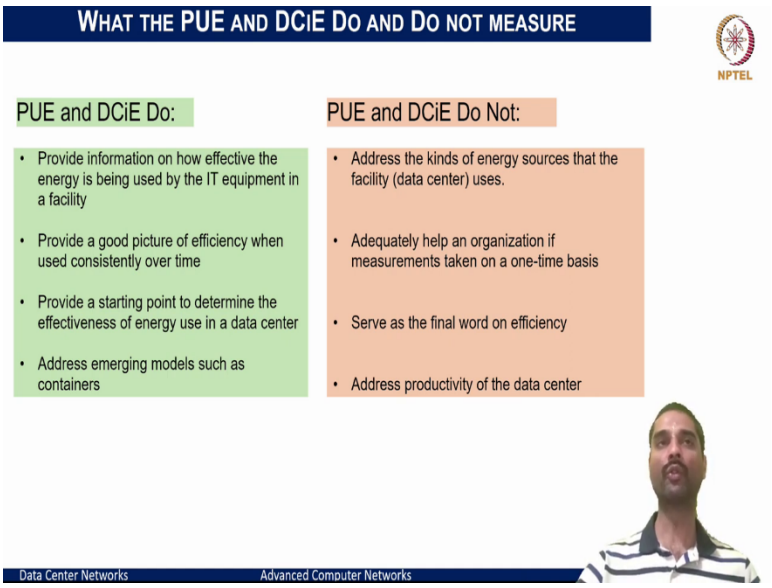
And while the PUE of 1.2, that is, 83% or higher DCIE, would mean highly efficient data centers, and what we saw right now with a value of 1.25, we are in the range of an efficient data center category, and this would build likewise with PUE of 2, 2.5 and the ratings and most of these metrics have been standardized across the globe in terms of how we would want to categorize the data centers.

And there was a recent study that was done by Uptime Institute in 2022, and it mentioned that on average if we look at the PUE ratings of all the 800-8000 data centers that we mentioned, it falls around 1.55. So, on average, we are somewhere having better or efficient data centers, but the

number of older data centers typically have a PUE rating of 2.5 to 3, and all the newer data centers fall within this bracket. And that is where our goal would be to bring this down further towards 1 or 1.2 so that we are able to have better energy-efficient data centers.

And there was lot of research driven by the industry as well as Academia towards how the Energy Efficiency of the data centers can be made up. And several of the things also came out in the light in terms of how the data centers can be deployed to be better, or energy efficiently done or reuse green energy as well. So, let us try to also look at what this PUE and DCIE tell us and what are the key other metrics that we would need to measure and quantify the Energy Efficiency and Green energy index of the data centers.

(Refer Slide Time: 11:51)



So, foremost what this PUE and DCIE do is to provide us the information on how effectively the energy is being used by the IT equipment in a data center facility. But note, that we are only talking about, I give so much of power and of this power a portion of it whatever is being apportioned for the IT equipment. But we are not justifying at all about whether this IT equipment's energy is really translated to effective work or is it just that I keep the fridge on at -20 degrees when I really do not need it to be on at all. But nonetheless, they provide a very good picture of efficiency when we use it consistently over time, that means, doing one measurement is not sufficient.

It is important to do this periodically, check what numbers we get, and it is also a thing to consider and see that the PUE index may vary over time because the IT equipment wore out and they may have different power requirements although the amount of power that you give the way the productivity of that device would change over time.

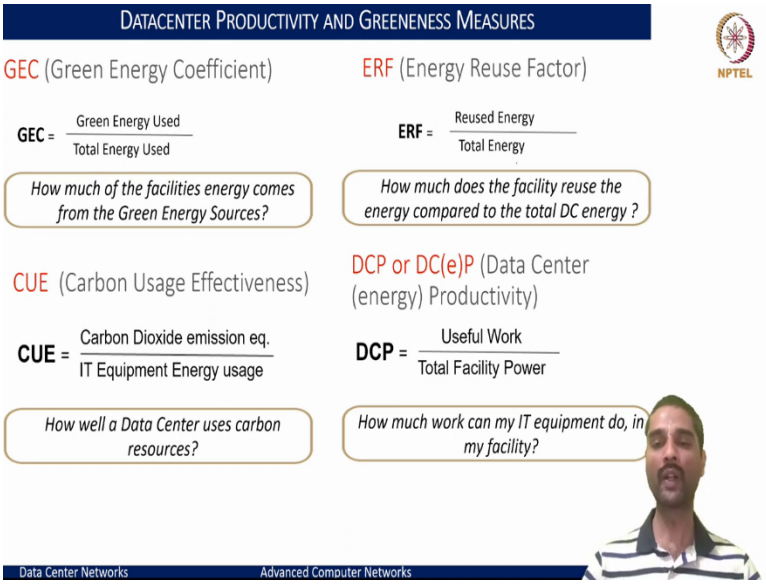
This is what we need to keep in mind when we do this, but also, there has been a good amount of debate in terms of where should you be measuring, right at the edge where the energy is being sent, at the rack level, within a data center should all be done and they were standardized over the years, and there have been a set of bodies that dictate about how the measurements should be done.

But we should also keep in mind that this PUE and DCIE does not address the kinds of energy sources that the facility data center uses; it is just the cumulative sum of what is the energy that is being used, regardless of what is the source that is actually driving the energy, because some part of it may come from internal reproduce, some part of it may be run from the grid and you may have a green energy that is being produced as well. So, the source of energy is not accounted when we look for the PUE and DCIE numbers.

Further, like they do not serve as a final word on efficiency because they only serve as how much is being taken out for IT equipment but whether that equipment is really doing any work or not is not being accounted, that means, these cannot be the terms for judging the efficiency of how the energy use happened.

And neither do they address anything about productivity because I give some amount of energy, I spent on it what is the revenue that you get back, or what is the way that that energy that you have spent is being utilized? None of these questions are being addressed by PUE and DCIE. And that is where there was a need to also bring in newer questions to define these particular characteristics.

(Refer Slide Time: 14:29)



And in this regard, Green Evolutionary Coefficient is the measure of how much of the facility's energy comes from green energy sources because we do not want the data centers to be driving a full load from energy sources that are not sustainable or have adverse impacts on the environment. And this is where it is important to account for how much of the energy of the total energy that we need comes from the green energy sources. And this metric, as we can see, is calculated by dividing the green energy used, the source of the data power that is consumed from the green energy sources over the total energy that we would want for the data center to operate.

And the closer this GEC is to 1, that is, 100 of the energy that is being treated in the data centers is run through the green energy, the better it is, or it indicates the best possible greenness quotient that we can define for a data center. But the green energy in itself has a lot of drawbacks because these are, in essence, sporadic and not consistent in terms of having the load-driving capability to be saved all the time. So, there are always some trade-offs to play with respect to how GEC can be used.

And the second of the metrics is the Energy Reuse Factor because, as we know, the data centers generate a lot of heat, and you also need a lot of cooling infrastructure to build around, but heat is, again, if you think is a kind of energy that is being generated from these data centers. So, how effectively can we reuse such energy for different work categories to put the heat into use for different kinds maybe you want to generate electricity itself so that whatever the heat that is being generated, if it can be channelized and used to regenerate electricity to power up the data

center again or account to a portion of the total energy being used by the data center come from this reused energy then that is a very important aspect.

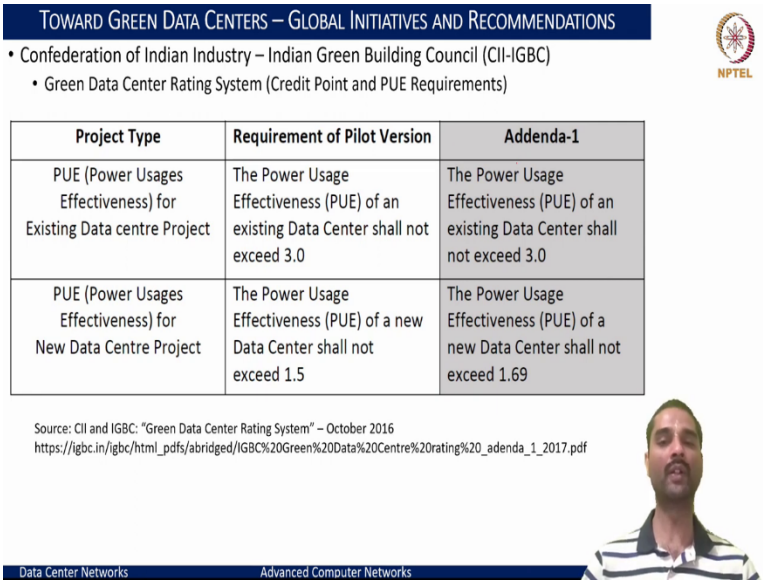
So, that is where this Energy Reuse Factor metric was defined, and this determines how much of the facility you use happens from the energy perspective compared to the total data center energy. So, it is basically the ratio of the Reused Energy over the Total Energy. And again, like GEC, we would want this ERF to be as close to 1, but it is very rarely possible that you can reach to 1, but at least the number, that is, ratio, that is somewhat close to 1 would be a better fit.

And overall, when we see that these data centers consume power, it could be from renewable or non-renewable sources, it is important to now understand what is the Carbon Quotient for these data centers, and that is where a new metric called CUE or Carbon Usage Effectiveness was defined. What this essentially tries to do is this metric determines how well a data center uses the carbon resources, that is, whatever the energy that we are going to be consuming from the non-renewable resources carbon emission and the amount of carbon emissions that is equivalent for driving such an IT equipment, over all the total IT equipment energy.

So, if you think of like if we have the GEC, 1 minus of this, in essence, would be similar to having the CUE, but this can be a lot more than this because whenever the energy is produced and heat is dissipated, this is also again going to add a more pile up for the Carbon Quotient and this is where we want this CUE to be as low as possible and a lower CUE would present a more optimal use of carbon for a given data center, eventually in terms of the productivity around 2008-2009 the Green Grid Forum introduced to the new Benchmark called the DCP or the Data Center Productivity, also a variant of it called the Data Center Energy Productivity specifically to measure or probe the useful work that is being produced by the data center for a given facility power.

Although, this term is somewhat ambiguous in the sense of how to measure it because how do we quantify the productivity is it in terms of the revenue is it in terms of some kind of computation work that is being done? So, what exactly constitutes to it I mean, they will try to define a lot more metrics, but having refined metrics to define useful work is a key here in trying to come up with the quotient of DCP.

(Refer Slide Time: 19:07)



And likewise in India, there was a lot of push and drive towards when the data center started to emerge, how to define and use these metrics, and what are the standards that we should be bringing. And this is where the confederation of Indian industry, the CII, and the Indian Green Building Council. So, together CII and IGBC came up with the initiative called the Green Data Center Rating System.

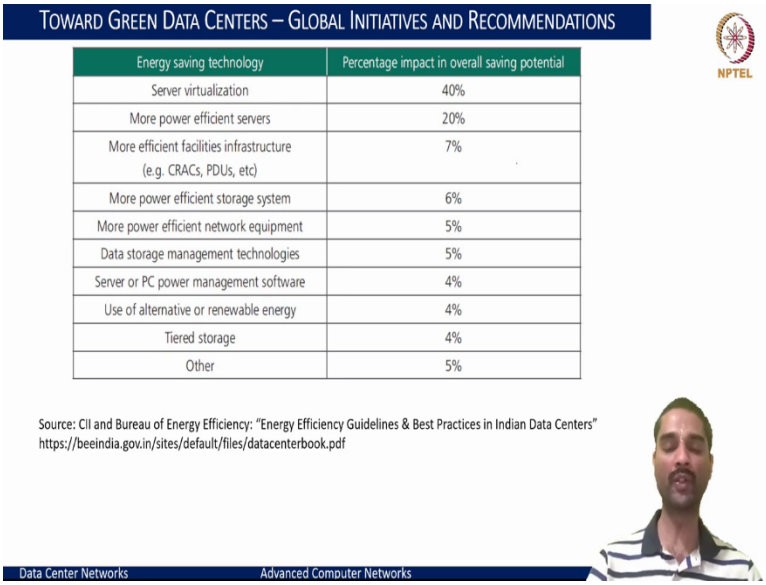
And this is a Credit Point based system based on the techniques the methodologies that have been adopted in the data center and what is the PUE index that you measure for the data centers over time and accordingly, a credit or a rating-based system was put out to say what is the green data center rating that we can for the data centers come in India. And all of this happened around October 2016, where the green data center rating system came.

And there were also some of the add-in terms the changes that were proposed on how the rating or metric should be and what we should be concerned about in this aspect is, they tried to say what should be the PUE numbers for the existing data centers and what we should be aiming for the PUE numbers for the upcoming data centers and in this regard they defined that the PUE for an existing data center should be around 3.0 and should not exceed 3.0 that means even if they are inefficient their PUE should not be more than 3 they have to be contained.

And for any of the new upcoming data centers, the initial code was to have the PUE less than 1.5, but then over in years time it was changed to say roughly around 1.69, and these are the

ratings that any data centers now if you have to come up in India have to try to meet towards these numbers.

(Refer Slide Time: 20:56)



And the report by the CII and the Bureau of Energy Efficiency also came out in terms of what should be the best practices for the Indian data centers and what Energy Efficiency aspects should we be looking for as we go and build these data centers. And this came up with a nice, rich set of guidelines to say what are all the practices that need to be done, and around 10 of the key prospects of how to build the Green Data Centers.

And here I am trying to cover some of the key works, key aspects that they try to show, and this is a table from the data center report which shows what should be the technology, that we should be employing to save more energy. And you can see the foremost is server virtualization, that means, we want to run and virtualize more and more of the servers that are going to run on physical devices wherein you consolidate a majority of the web workloads or any other kinds of workloads onto a single physical machine and again we can see that this is where the software comes to picture in terms of how we virtualization and softwareization of the entire networking components when we spoke of the network function virtualization, when we spoke of the network virtualization, when we spoke of the containers, the server had less workloads, all of these are the important prospects in this shape of the serverless, server virtualization that would

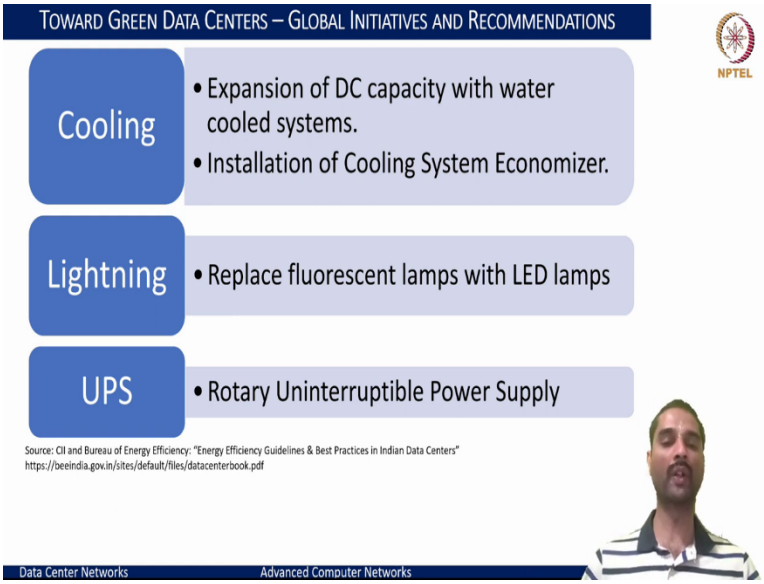
account to the majority of power savings and it is estimated that more than 40% of the power savings the impact would be from this virtualization.

And also, we would want to use more power-efficient servers and more devices that are basically newer technologies where you are using less of a power to drive those devices than what used to be in the earlier case, and that would also account to more than 20% impact on the overall energy.

And the third aspect is the cooling infrastructure, how you would want to build the CRACs, the PDUS all of these for infrastructure ways of cooling how you would want to build, and even the power loss we want to contain, and we saw earlier that 21% accounted to the power losses. So, how do you efficiently store the power and reuse that power becomes a crucial aspect, and mechanisms here would also drive towards a good contribution towards the overall savings on energy.

So, likewise, the other use of the alternative or renewable energies, how to better build the power management software within a system so that you are able to basically scale the device processing speeds based on the workloads and based on the power and heat aspects, all of these have to be pulled in. And now, we can see that most of our devices have this scaling of the processor's frequency to adapt to the workload and cut down on the power requirements when you do not have any work to run. So, all of these have already been in place in most of the devices.

(Refer Slide Time: 23:49)

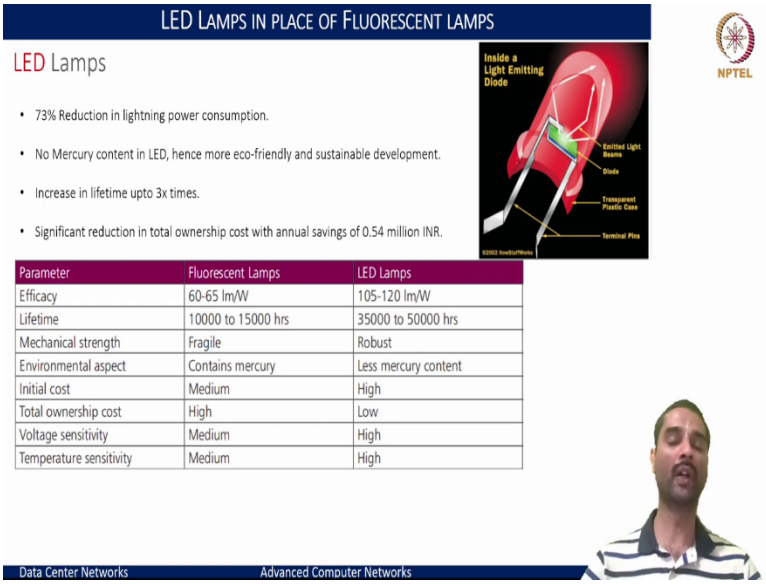


So, to quickly look at a couple of the aspects, one is in terms of the cooling the data center report comes up with saying you want to have the expansion of the data center capacity with water-cooled system. So, it was seen that having the water-cooled systems, in a way, saves a lot more energy than running it through the air conditioners or other models.

And such an installation of a cooling system economizer which keeps track of what are the modes to cool, when do you want to build, like which area you want to focus on within a big data center, where you want to cool like we initially spoke about the hot aisle and cold aisle maintaining and ensuring the air circulation within those hot aisle and cold aisle helps significantly improve including the raised floors that we spoke of and with the ventilation for circulation of the air.

All of these infrastructural and Engineering aspects of how you build the data center contribute a lot towards saving on cooling energy. And second, with respect to lightning because if you are hosting so many of the data centers, you would want the light to be there to be having to look at and run them continuously and one of the very simple but most efficient mechanism that this project even showed is replacing fluorescent lamps with LED saves or contributes to a significant savings on the energy and likewise with respect to the uninterrupted power supply that you use using a model called a Rotary UPS helped significantly save the energy.

(Refer Slide Time: 25:14)

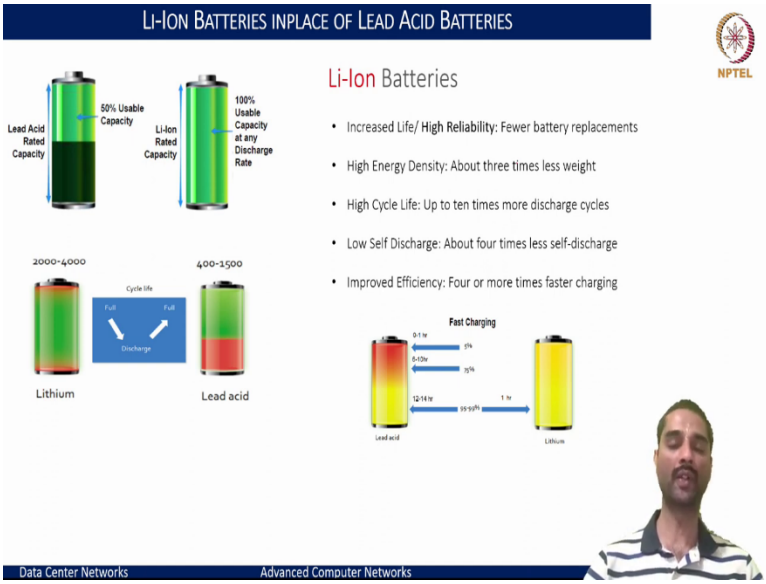


So let us look at basically one of the prospects when it comes to lightning what this project from DBE also tried to see is if you just replaced the fluorescent lamps with LED lamps, there were a lot of benefits that you would get, not just in terms of power but also in terms of the overall capacity and cost plannings and even making it more sustainable and eco-friendly development. Like if we say fluorescent lamps were basically 60 to 65 lumens per watt while LED lamps are able to provide 105-220, almost double the Lumen per watt, and this is where the efficiency comes.

And even although the investment initial cost is high, the lifetime for these is almost 2 to 3 times more than the lifetime of LED lamps, and that is where the initial cost that you put for LED lamps would eventually pay out over a longer run. And these are being more robust, the replacements would be much lesser the operational and management costs could also come down and that is where this project, which was run for over two years, showed that using just the LEDs helps reduce almost 73% in the lightning power consumption.

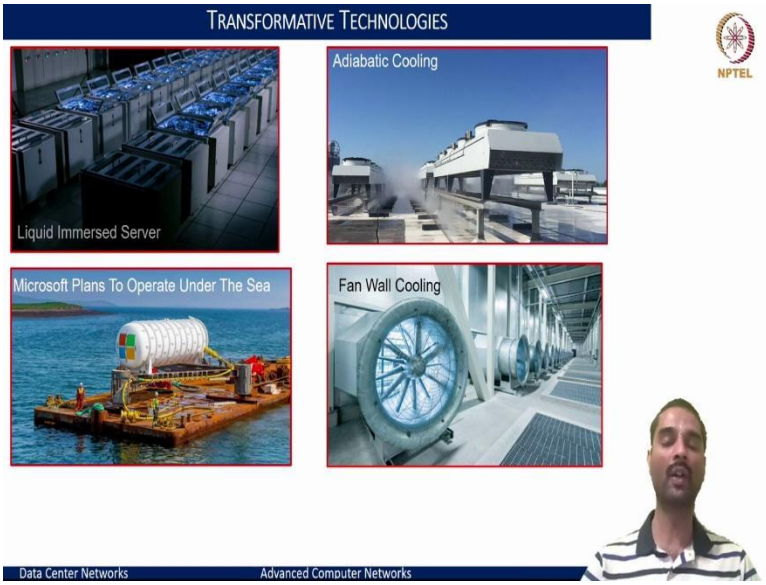
Although the overall power consumption of lightning maybe 1 to 5% in a data center but if you are able to save 73% on that in reducing it is a significant portion, and that is what it eventually said a total ownership cost was brought down to 0.54 billion INR in just over 2 years while providing the lifetime of these LEDs to be 3 times more than that of the fluorescent lamps.

(Refer Slide Time: 26:49)



And likewise, when we speak of the batteries that we use to drive the power, the Global Green Grade Consortium also presented a point towards replacing the lead acid batteries, which will be earlier era batteries to the Li-Ion batteries which would likewise provide increased or high reliability, providing better energy density, and able to have better capabilities in terms of charging and discharging power with higher number of life cycles of charge and discharge, and also being more eco-friendly, because now there is not much of an asset aspects that you want to see where you would want to dispose them. And this is where the aspect of how to switch over to the Li-ion batteries was also being discussed and being made a case to be used in data centers.

(Refer Slide Time: 27:39)



And there were several other transformative technologies that have come around in terms of how you would want to build the data centers themselves. And one of them, beside the servers, generate a lot of heat. So, the novel technology that came around is to have a liquid immersed server that means you submerge the server into a bath of a non-conductive liquid and allow the thermal energy that is generated by these components to be soaked into that fluid and automatically make the systems get cooled.

And this immersion cooling has almost now been a standard in terms of how modern data centers are going to be built, which will have a very low footprint on carbon emission and very low overheads on the water usage as well needed for cooling, and it is been seen as a more scalable faster and energy efficient means to dissipate the heat.

And likewise, to the liquid emerged servers, the other aspect is when and how do you cool this data center system, and in that prospect, the adiabatic cooling scheme will, unlike the traditional cooling towers that continuously keep circulating the air and water to cool this infrastructure, you would want to do it in a more random way and do it using the evaporative cooling scheme. And what this means is they would run dry for an almost very large portion of the time and only use the cooling infrastructure as necessary, wherein the ambient air that is hot enough could also be used as a means to circulate the air well and cool the system.

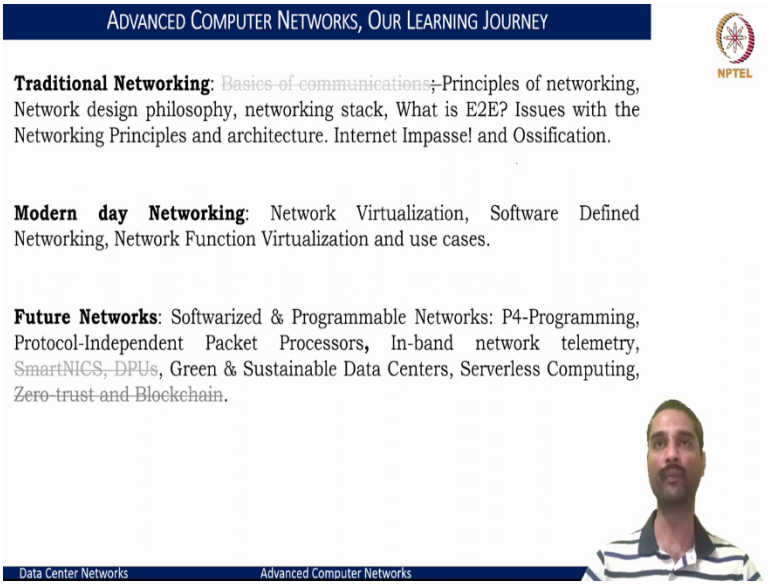
And in fact, Microsoft tried to bring in the concept why cool when you have these ocean beds; why not deploy the data centers under the sea? And this was a project that Microsoft started called I think, the project Natick around 2018, and they ran it successfully for over 2-3 years to see what happens when we run the data center under sea, and the picture here is the barrel that we see was actually being submerged to keep it in the sea to see how it would operate.

And it gave a lot of interesting results in terms of how things would perform, what is the reliability prospects and it was seen that this operated much more efficiently. And in the future maybe we will see a lot more of the undersea data centers come up, and then there will be a lot more aspects to think, how the things would be taken, how do you set up the infrastructure, although the initial costs may seem high but nonetheless these pay different options.

And likewise, the fan wall-mounted cooling has also been seen as a better mechanism to dissipate the heat where the wall cooling is are being set up at the data centers, and the raised

flows would be continuously driven by these fan wall coolers. And this area has seen a lot of attraction, and there is a lot of research not just from the computer science side but from all the mechanical, civil engineering operations in terms of how the data centers need to be built, the technologies in terms of what electrical components you need to use, all of these prospects have opened up when we think of the data centers and several of the research is being going forward in this area.

(Refer Slide Time: 30:46)



So, in a nutshell, we have tried to cover a lot of aspects in the last five weeks. So, we looked at some of the key Advanced Topics in the data center networks in the overall networking community, starting from the traditional networks to modern-day networking, and looked at the key areas that have come up in the last 10 to 15 years. We were able to shed some light on these developments in virtualization, software-defined Networks, NFV serverless Computing, the cloud models of services, and I hope you have all enjoyed this part of the course.

And if you find any of these areas of interest store, you want to discuss further on any of these aspects do write back to me at my email id sameergk@iitgn.ac.in , and next up Professor Neminath will be addressing on another important aspect called the Future Internet, and this would cover basically the aspects of how the networks are becoming more and more information-centric and what it means to drive the network not as a means to the communication of the data or reach to other ends through an IP, but use the different prospects of how the

information can be seen as a network itself and you want to build the network above around this information and this is where the ICN, Information Centric Networking and Named Data Networking or NDN aspects will be covered. Thank you.