Institute of Architecture of Application Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Masterarbeit

Data Center Simulator for Sustainable Data Centers

Emad Omar.

Course of Study: Computer Science - STE

Examiner: Prof. Dr. Marco Aiello

Supervisor: Brian Setz, M.Sc.

Commenced: April 1, 2019

Completed: October 1, 2019

Acknowledgement

First of all, I would like to express my gratitude to Professor Marco Aiello for giving me the opportunity to work on this thesis at the Institute of Architecture of Application Systems, University of Stuttgart.

I would also wish to express my gratitude to Professor Frank Leymann. He motivated me by his unique character and teaching style during his valuable courses in the master program.

Further, I would like to express my special appreciation to my master thesis supervisor M.Sc. Brian Setz for his continuous support, fruitful suggestions, constructive comments, patience, and insightful guidance throughout all stages of this work, without him this thesis would never have been completed.

I would like to express my profound gratitude to my mother, my father, my brothers, and my sister for their love and constant encouragement throughout my life.

I extend my sincere gratitude to Hamdy, Amr, Mogy, Ahmed and Alaa. I greatly appreciate their support, motivation and love.

Finally, I would like to thank my girlfriend. Her unconditional love, sense of humour, patience, and support was more valuable than she could ever imagine.

Abstract

Current Data Center (DC) simulators focus more on resource management and deployment of applications, including the new virtualization technologies that can be used to simplify the deployment and improve resource efficiency. While this is an important perspective for the Development and Operations (DevOps) engineers, DC operators on the other hand have different priorities, especially the infrastructure challenges such as cooling and power consumption, and how changes in the workload will impact the energy efficiency of a DC. A DC simulator, which can consider the environmental impact, and evaluate the different sustainability metrics, can lead to more energy efficient and sustainable DCs in the future. Hence the main target of this research is to study the different sustainability metrics and their input parameters to determine the minimum number of commonly reused inputs that are expected to evaluate the maximum number of sustainability metrics, which consequently helps in designing more energy efficient and green DCs. This can be done by selecting the commonly reused set of inputs and metrics, and researching the parameters that are essential to assess and evaluate the sustainability metrics of a DC. It requires a detailed study and profound elaboration and analysis of the set of metrics and inputs including their interrelationships, and experimenting how changes in the DC parameters are reflected in the metrics. There are many metrics classifications available to evaluate the different key performance indicators. A collective and integrated classification of these sustainability metrics is provided in [VSR+17]. The paper classifies the metrics into 9 categories, starting with the energy efficiency metrics down to the financial metrics, and it gives the mathematical equations required for calculating the different sustainability metric.

Contents

List of Figures

List of Tables

Acronyms

CFD Computational Fluid Dynamics. 68

CO2 Carbon Dioxide. 7

CPU Central Processing Unit. 27

CRAC Computer Room Air Conditioning. 44

CUE Carbon Usage Effectiveness. 75

DC Data Center. 4

DCeP Data Center Energy Productivity. 77

DCiE Data Center Infrastructure Efficiency. 34

DCP Data Center Productivity. 70

DCPE Data Center Performance Efficiency. 77

DevOps Development and Operations. 4

EDC DC Energy Consumption. 7

EPi Planned Energy. 34

ERE Energy Reuse Effectiveness. 82

ERF Energy Reuse Factor. 39

EWR Energy Wasted Ratio. 33

GEC Green Energy Coefficient. 36

GHG Greenhouse gas. 14

GUI Graphical User Interface. 66

H-POM IT Hardware Power Overhead Multiplier. 33

HVAC heating, ventilation, and air conditioning. 37

IaaS Infrastructure as a service. 27

IT Information Technology. 13

ITEE IT Energy Efficiency. 34

ITPUE IT Power Usage Effectiveness. 75

ITUE IT Usage Effectiveness. 34

PUE	Power Usage Effectiveness. 7
RAM	Random Access Memory. 66
SLA	Service Level Agreement. 63
TCO	Total Cost of Ownership. 15
UPS	Uninterruptible Power Supply. 7
VM	Virtual Machine. 27

1 Introduction

In a world with a growing population and decreasing resources, information technology is considered to be a driving force for changes all over the world, therefore computer scientists and people who are involved in the information technology industry must take the responsibility to make the world more greener, and try to optimize the utilization of resources while trying to minimize the dangerous environmental impacts which result from the computational infrastructures. Furthermore, the increasing demand for cloud-scale services such as storage, networking, and computation has driven intense growth of large complex DCs that run many of today's web, financial and business applications.

DC has become an increasingly important part of most business operations nowadays, and as a result there is a continued rise in electricity that is consumed by DCs. This increase in power consumption resulted from a number of different factors like: more computers, laptops and internet-enabled mobile phones, similarly more powerful applications and a demand for faster internet speeds, in addition to more internet-based business, which consequently causes increased internet-access and increased numbers of businesses, besides more intensive online activities such as social networks and media sharing sites. Furthermore, at the server level, chips have become increasingly smaller as transistors have reduced in size, and although there has been an increase in computations per kilowatt hour, performance increases at the same time as server density in DCs, which accordingly leads to higher demands on power and cooling [KBSW10].

The energy use and environmental impact of DCs has recently become a significant issue for both DC operators and policy makers. The international awareness of climate change and it's disastrous impacts on the environment has changed significantly, delivering real commercial impacts for corporate environmental policy and social responsibility. DCs represent a relatively easy target due to the very high density of energy consumption and ease of measurement in comparison to other, possibly more significant areas of Information Technology (IT) energy use. Policy makers have identified IT and specifically DC energy consumption as one of the fastest rising sectors. Simultaneously, the commodity price of energy has risen faster than many expectations. This rapid rise in energy cost has substantially impacted the business models for many DC operators and has already driven changes in the way DC capacity is charged for commercially. Energy security and availability is also fast becoming an issue for DC operators as the combined pressures of fossil fuel availability and environmental energy policies make it difficult to predict energy availability and cost. Figure 1.1 shows how DC operators are trapped between demands and constrains, while trying to achieve conflicted requirements related to carbon footprint [New08].

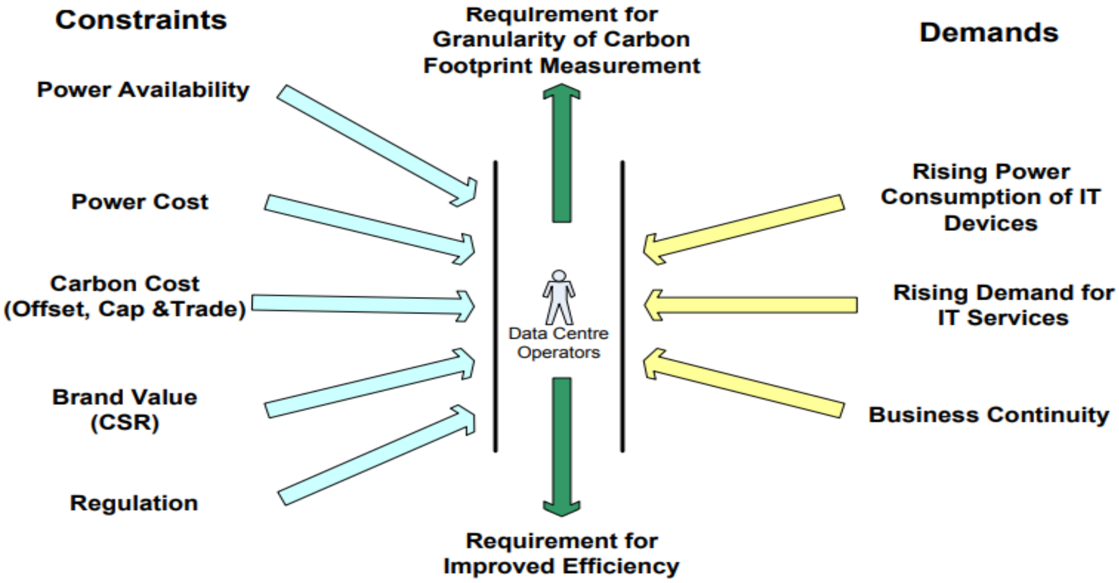


Figure 1.1: DC operators trapped between user demands and energy constraints [New08]

There are many reasons that make DCs have been identified as one of the fastest growing consumers of energy. According to the United States Environmental Protection Agency (EPA) report, US DCs consumed 61 billion kilowatt-hours of power in 2006. That represents 1.5% of all power consumed in the United States and represents a cost of $4.5 billion. Furthermore, In 2007, DCs in Western Europe consumed 56 terawatt-hours(TWh) of power per year. According to the EU, this figure is likely to almost double to 104 TWh by 2020. This projected growth, if not offset by innovations in efficient energy management, will prevent the European Union from achieving its overall carbon reduction and climate change targets. Figure 1.2 shows the growth of worldwide atmospheric CO2 [Bou11].

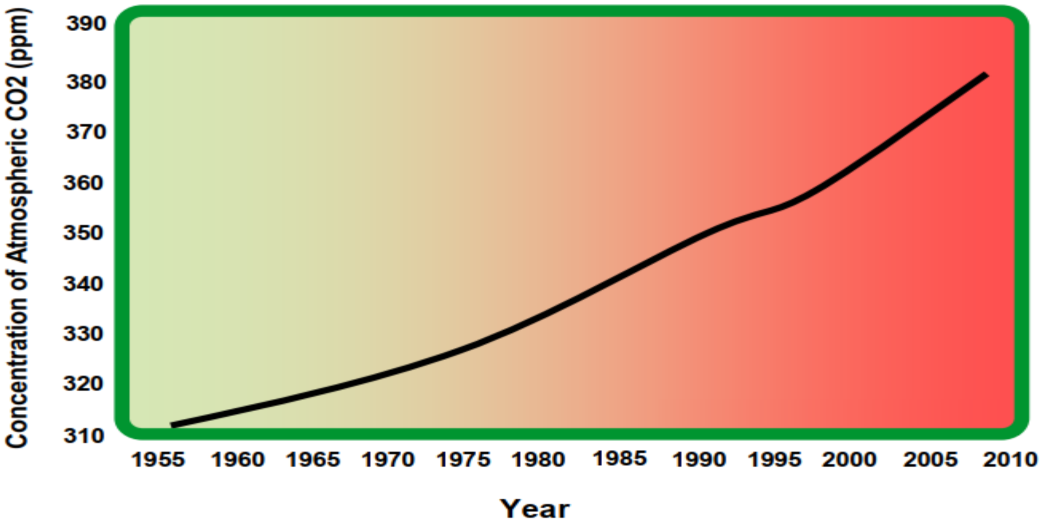


Figure 1.2: Growth of worldwide atmospheric CO2 [Bou11]

DCs require large quantities of electric power that meets better quality and availability levels. Therefore, huge capital investments are required to provide power, air conditioning and ventilation. Moreover, the increasing electric loads led directly to forecast increases in energy-related air pollution as shown in Figure 1.2. Greenhouse gas (GHG) emissions such as carbon dioxide from fossil fuel-fired electric generators, are a major cause of global climate change. The forecast increases in DC electric demand implied a growing source of GHG emissions. Consequently, the international community is attempting to reduce GHG emissions amid broad scientific and international consensus that climate change is a major threat and must be taken seriously. Air pollution from diesel generators causes environmental health problems. Each planned DC includes as many as two or more times redundant backup diesel generator capacity, to use when the electric grid is unstable or unavailable. Diesel generators used for emergency backup power supply are essentially unregulated. They are a notorious source of very high levels of damaging air pollutants, including soot, nitrogen oxides (NOx), which form ground level ozone (smog) during hot sunny weather. Because of concerns over pollution and strains to electric utility infrastructure, many organizations and academic researchers have begun to examine how to improve energy performance in DCs, so as to reduce energy use and consequent air pollution emissions [Rom19].

Sustainability as an infrastructure strategy can be defined according to the following principles such that, any energy consumption should be kept as low as possible and any resource should be used as fully and efficiently as possible, which means that wastage in resources should be minimised. Moreover, timely and accurate information should be produced to assess energy usage, efficiencies and resource use. Additionally, the full environmental and social impact of activities should be considered, and the level of IT resource provision should be adjusted to the task being done [Gel].

DCs that have a sustainability emphasis on energy consumption, carbon emissions, and water usage have more control over their decisions on growth, location, and outsourcing strategies. With more sustainable DCs, IT organizations can better manage increased computing, network, and storage demands and at the same time lower their energy costs and hence reduce the the Total Cost of Ownership (TCO) of its IT infrastructure. TCO quantifies the financial impact of deploying an IT product over its life cycle and includes its energy cost, which is usually the biggest factor and additionally other costs like installation, licensing, maintenance and support. The future presents risks, especially when it comes to carbon taxation, water costs and policies. Organizations that proactively focus on these issues will lower their business risks, increase their potential for growth, and better manage their environmental costs [Gel].

1.1 Problem Statement

Nowadays, the rapid growth of information leads to design challenges in DCs which have to be able to adapt to the constantly changing requirements resulted from the growth in volumes of data. Consequently, DC operators have to deal with the scalability issue and the environmental impacts which are resulted. Therefore, determining the impact of changes in workload and how the consequences affect the sustainability of a DC, are priorities for DC operators. DC operators need for improved sustainability, because DC power footprint, energy usage, and carbon emissions are affecting companies' decisions on growth and building locations [APPT10].

It has been reported that, in year 2006 alone, the DCs and servers in U.S. consumed 61 billion kilowatt hours of electricity, which is 1.5% of all US electricity consumption and has a monetary cost of $4.5 billion [Bro+08]. These massive amounts of computation power that are required to drive and run these server farms cause many issues, like huge energy consumptions, emission of greenhouse gases. Therefore, energy efficiency has emerged as one of the most important design requirements for modern computing systems, such as DCs, as they continue to consume enormous amounts of electrical power and result emissions of carbon dioxide into the environment. Currently IT infrastructures contribute about 2 % of total CO2 footprints. Unless energy-efficient techniques to manage computing resources are developed, It's contribution in the world's energy consumption and CO2 emissions is expected to rapidly grow [UA12].

Environmentally sustainable DCs are built to consume the least amount of the most appropriate sources of energy throughout their lifetime [BPSS08]. These sustainable DCs intended to reduce the dangerous environmental impacts through achieving the balance between the need for increased performance, and reducing energy consumption.

In fact, there is a large number of inputs that are used for evaluating the different sustainability metrics and measuring the performance of a DC [VSR+17]. Therefore, in this research, all of the inputs and metrics are analysed and the interrelationships between them are illustrated by using tree graphs, in order to limit the large number of inputs to a restricted set of commonly reused inputs. Hence, DC operators can evaluate some of the important sustainability metrics using a limited number of commonly reused inputs.

1.2 Research Questions

The simulation of DC is done after investigating and analyzing the whole set of inputs and the outputs of the sustainability metrics, and identifying the set of commonly reused inputs and metrics. This DC simulator is intended to show how can DC operators achieve sustainability based on using this subset of commonly reused inputs. In order to do this, the following research questions will be addressed throughout this work:

- **RQ 1**: What are the commonly reused inputs and metrics in a DC?
- **RQ 2**: How can we use the commonly reused inputs and metrics to simulate a sustainable DC?
- **RQ 3**: How can this DC simulator help DC operators to achieve sustainability?

1.3 Thesis Organization

The thesis starts with background research in Chapter 2, in which we provided a general introduction about DCs and the main components of a DC. Then a discussion about DC issues that are related to power consumption and environmental impacts is given. Finally, this chapter ends with the state of the art and recent trends related to DC sustainability. In chapter 3, The related work is provided, and followed by the contribution of this thesis. The core of this research is the sustainability metrics analysis, which is provided in chapter 5. In this chapter, all the DC input parameters and metrics are analysed, and a tree graph is created for each category of sustainability metrics, these tree graphs illustrate how the metrics are calculated using the different inputs. We provided a table that shows the commonly reused inputs and their frequencies, and then a rooted tree graph is created for each commonly reused input. By investigating these commonly reused inputs, we found that 16 of the sustainability metrics can be evaluated using 15 commonly reused inputs. In chapter 6, a set of the existing DC simulators are analysed to check if it is possible to use any of these simulation platforms to simulate the commonly reused inputs, but the comparison shows that, there is no simulator that can fulfill our requirements. Therefore, the sustainable DC simulator is presented in order to evaluate the maximum number of metrics based on using the minimum number of commonly reused inputs. Using the sustainable DC simulator, 2 simulations for 2 samples of inputs are are done and results are provided. The simulation results are visualised in chapter 7, in which we provide a discussion about the results of the metrics analysis and the sustainable simulator in relation to the research questions. Chapter 8 gives a summary of the most important findings of this work, and it provides the limitations that restricted our experiments. Finally, this chapter ends with recommendations for future work.

2 Background

This chapter consists of three sections, in the first section, a general discussion about DC is provided. In the second section, a discussion about DC issues that are related to power consumption and environmental impacts is given. The third section is about the state of the art and recent trends related to DC sustainability including the most important sustainability metrics.

2.1 Data Centers

A single DC can house many thousands of servers and can use as much energy as small city [UA12]. Figure 2.1 shows a sample DC model, DCs house the computers, switches, routers, data storage devices and related equipment used to operate the digital economy. They can be relatively small facilities owned by a single company and dedicated to processing its data alone. At the other end of the scale, they can be huge facilities where many companies can rent floor space containing wire cages filled with racks to house their servers and other equipment [Rom19].

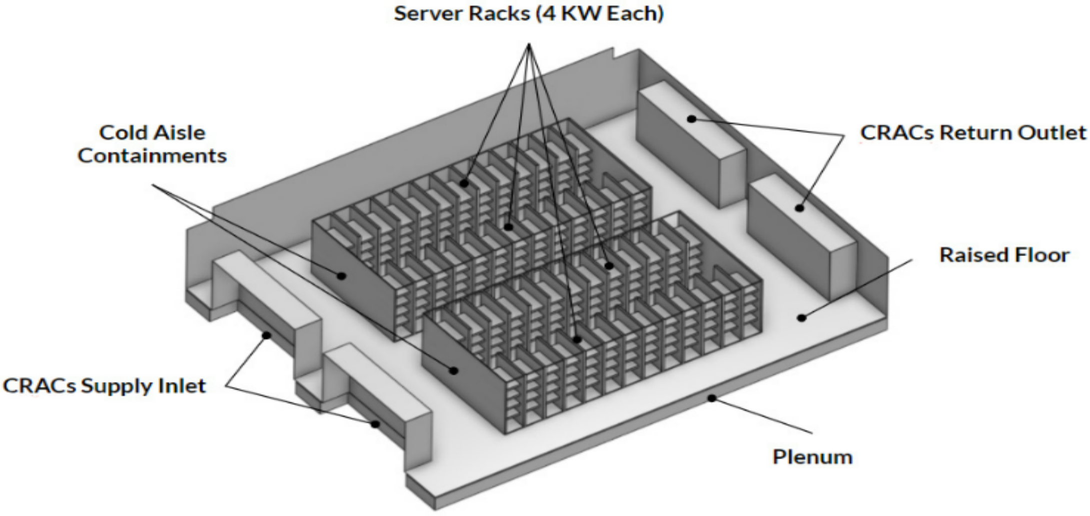


Figure 2.1: Sample DC model [Sima]

DCs are energy-hungry infrastructures that operate round the clock. They provide computing functions that are important for the daily operations of top economic, scientific, and technological organizations around the world. Figure 2.2 shows that, it is estimated that DC electricity consumption in US will increase to roughly 140 billion kilowatt-hours annually by 2020, corresponding to about 50 large power plants, with annual carbon emissions of nearly 150 million metric tons [CdL+18].

The financial impact for DC management is also huge, since a DC spends between 30% and 50% of its operational expenditure in electricity: the expected figure for the sector in 2020 is $13 billion per year of electricity bills (updated information can be found on the web portal of the U.S. National Resources Defense Council , The efficient utilization of resources in the DCs is therefore essential to reduce costs, energy consumption, carbon emissions and also to ensure that the quality of service experienced by users is adequate and adherent to the stipulated service level agreements. Efficiency is essential not only in single data centers, but also in geographically-distributed data centers, whose adoption is rapidly increasing [CdL+18].

The increased amount of energy consumed by DCs, which has a tremendous economic, environmental, and performance impact that makes the energy efficiency of cooling systems one of the primary concerns for DC designers, ahead of the traditional considerations of availability and security. Planning of the DC infrastructure design is critical, and scalability needs to be carefully considered. Another important aspect of the DC design is flexibility and elasticity, which are essential for hosting new additional services. Designing a flexible architecture that has the capability of supporting new applications in a short time, can lead to significant advantages [VLL+14].

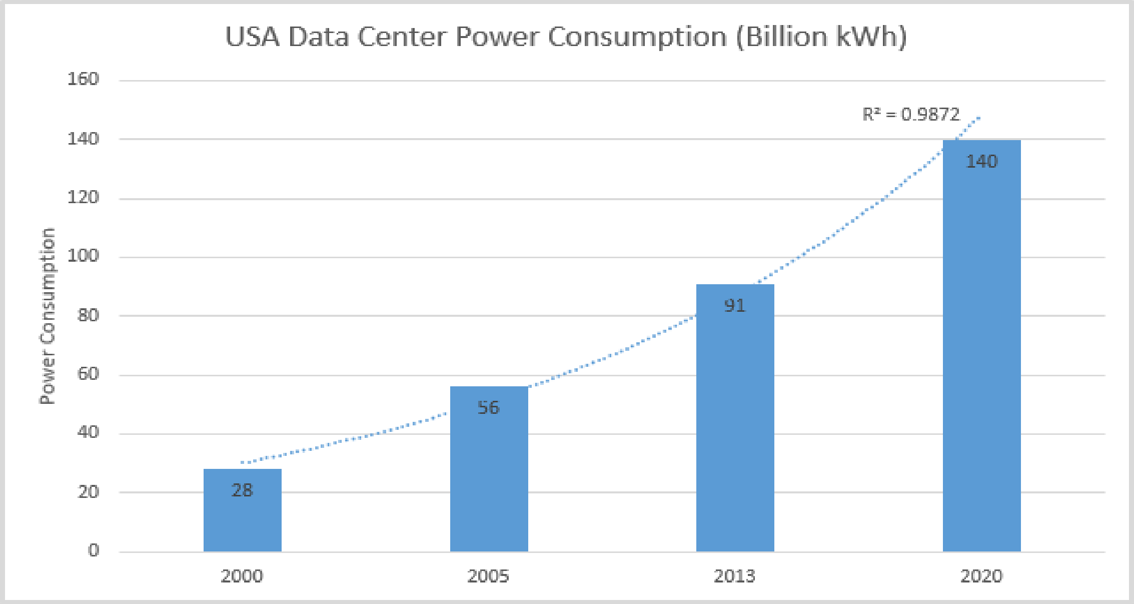


Figure 2.2: US DCs power consumption over years [Mod]

Data centers can be principally classified into 3 types according to the survey in [VKSK17], traditional, modular and green DCs. Traditional DCs are held and operated in fixed buildings and take years to construct. So it is unable to deploy and satisfy the flexible business requirements. Modular DCs, on the other hand, are kept in shipping-based containers, which includes UPS, Servers, cooling systems, storage devices and network devices in racks. They are easily deployed to meet the customer requirements. It can support more servers than traditional DCs.

But on the other hand, one of it's disadvantages is server compatibility. Additionally, modifications are difficult to be handled as they are produced by various vendors and equipped with exclusive servers. The third and most environment-friendly type is the green DC which uses energy-saving technologies, modular design and advanced power unit. Power Usage Effectiveness is an essential metric in evaluating energy efficiency. Most of the DCs start to use renewable energy resources to eliminate carbon emissions and operating costs.

DC facilities rarely achieve the operational and capacity requirements specified in their initial designs. The application of new technologies, such as blade servers, that require incremental power and more cooling capacities, as well as the pressure to house multiple DCs into fewer locations, and the need for incremental space and potential changes in safety and security regulations converge to impose constant facilities changes on the modern DC. The ultimate rule in DC facilities is to design for flexibility and scalability. This rule embraces several key principles in the site location, building selection, floor layout, electrical system design, mechanical design, and the concept of modularity that enables the DC facility to change and adapt as needed, with minimum renovation and change to basic building systems [Bel05]. Figure 2.3 shows the main components of a DC.

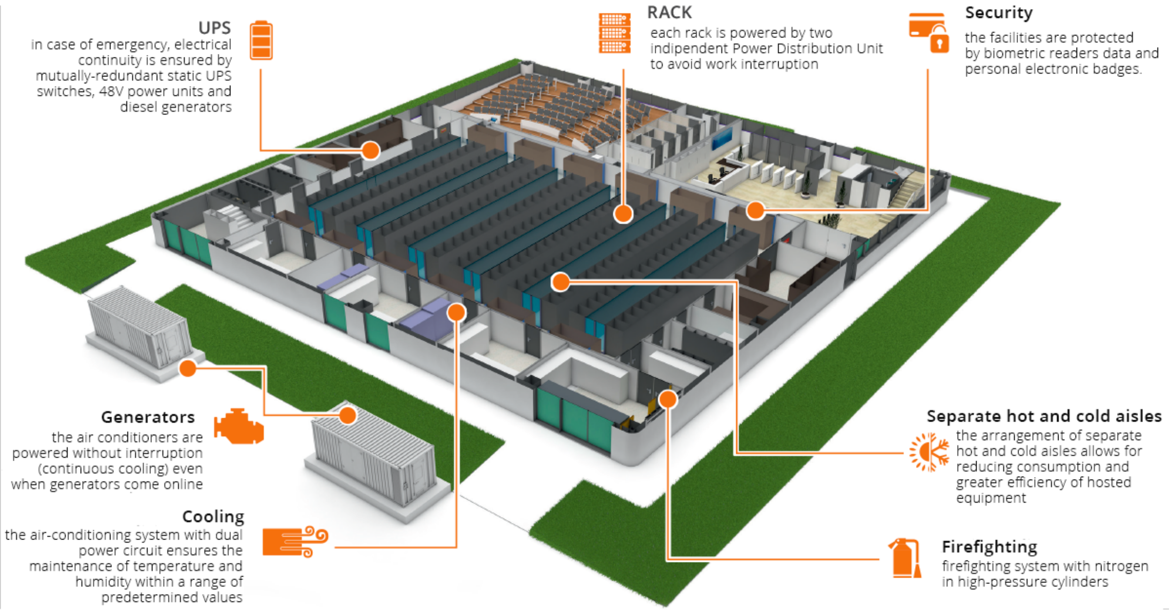


Figure 2.3: Main components of a DC [Wel]

2.2 Data Center Issues

Data centers are recently associated with larger data volumes, high energy consumption involving IT, power supply, ventilation and cooling. DC energy efficiency is a major concern for DC design and operation, because energy efficient DC means more green, environment-friendly and cost-effective DC. To improve DC energy efficiency through efficient cooling and ventilation, advanced and optimized processing and process models are required. Additionally, the accelerated pace of technologies such as mobile computing, Internet of Things, and cloud computing leads to a significant demand in the DCs, which in turn requires substantial incremental power and cooling capacity.

IT loads can go up (an increase in processing requirements from the lines of business) or down (impact of virtualization or consolidation). The higher the load, the more power will be required to keep it up and running and the higher the carbon footprint. Unluckily, the traditional practice in DCs of oversizing the physical infrastructure to fit the IT load has a very negative impact on DC energy efficiency and therefore, impacts carbon footprint. If a DC's electrical consumption is reduced, high carbon plant contribution will be reduced, which reduces the carbon footprint.

For the reason of the increased worldwide atmospheric carbon emissions illustrated in Figure 2.4, the U.S. Environmental Protection Agency cites DCs as a major source of energy consumption in the United States. Similarly, European Union (EU) members have agreed to cut their combined emissions of greenhouse gases to 8 percent below the 1990 level by 2012. Consequently, DC owners will be increasingly challenged to report their carbon emissions. A simple approach for estimating the carbon footprint of DC anywhere in the world was provided in [Bou11].

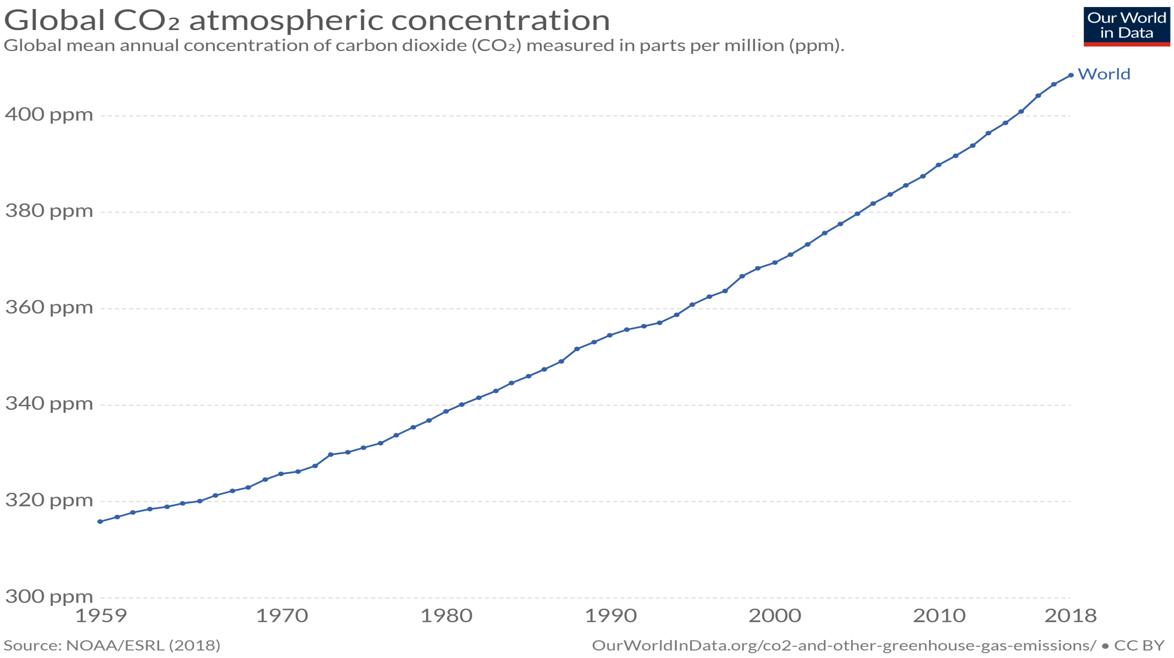


Figure 2.4: CO2 concentration [RR17]

A discussion about the three key factors that affect the carbon footprint of a DC was provided in [Bou11]. The three key factors are location, IT load and electrical efficiency. The geographical location of a DC plays an important role. A DC located in an area with easy access to hydro, nuclear, or wind power, for example, would have a lower carbon footprint than a DC located in an area that depends more heavily on coal, oil, or natural gas. The IT load consists of all of the IT hardware components that make up the IT business architecture: servers, routers, computers, storage devices, telecommunications equipment, as well as the security systems, fire and monitoring systems that protect them.

2.3 Data Center Sustainability

The accelerated growth in DCs energy consumption, which produces large carbon dioxide emissions as shown on Figure 2.3. Moreover, this growth requires higher electricity budgets. Environmental issues and energy prices forced companies to think about green DCs, which uses renewable energy sources, such as solar and wind. Measuring how resources are used in a DC is necessary to understand the overall efficiency, reduce the costs of operations and achieve sustainability goals. For efficient and environment-friendly operation of DCs, in [VSR+17], the paper monitors all the components of a DC and precisely classifying the components of a DC enabled them to assign a category to each of the DC sustainability metrics, which are considered as key performance indicators that are important for planning, designing, building and operating a DC in an efficient manner. The paper provided a survey of DC sustainability metrics with dimensions such as Energy Efficiency, Cooling, Greenness, Performance, Thermal and Air management, Network, Storage, Security, and Financial impact. For each metric, the paper describes the unit in which it is expressed, the objective, the optimal value as well as the scale at which the metric operates.

Guidelines for determining the PUE of a dedicated DC were provided in [AAFP12]. PUE is an excellent metric for understanding how well a DC is delivering energy to its information technology equipment. The PUE metric is associated with the DC infrastructural design. It measures the relationship between the total facility energy consumed and the IT equipment energy consumed. It provides strong guidance and useful insight into the design of efficient power and cooling architectures which helps in designing sustainable DCs.

Figure 2.5 provides a comparison between the percentage of carbon dioxide emissions resulted from DCs and and other industries, it shows DCs contribute with 0.2% of the total amount of CO2 emissions worldwide. Sustainability key issues such as carbon emissions and energy consumption in DCs, lead to research on energy savings. One of the most significant concerns in energy savings is smart cooling. High performance equipment requires more power which generates more heat and requires more dynamic cooling. In addition to the optimization of cooling resources through computational fluid dynamics modeling. [PBS+03] provides additional energy savings by manipulation of flexible cooling infrastructure to dynamically provision cooling resources based on demand. The paper uses real-time feedback from a rack-level temperature sensor network that is distributed throughout a DC to derive the cooling demand in different parts of the DC.

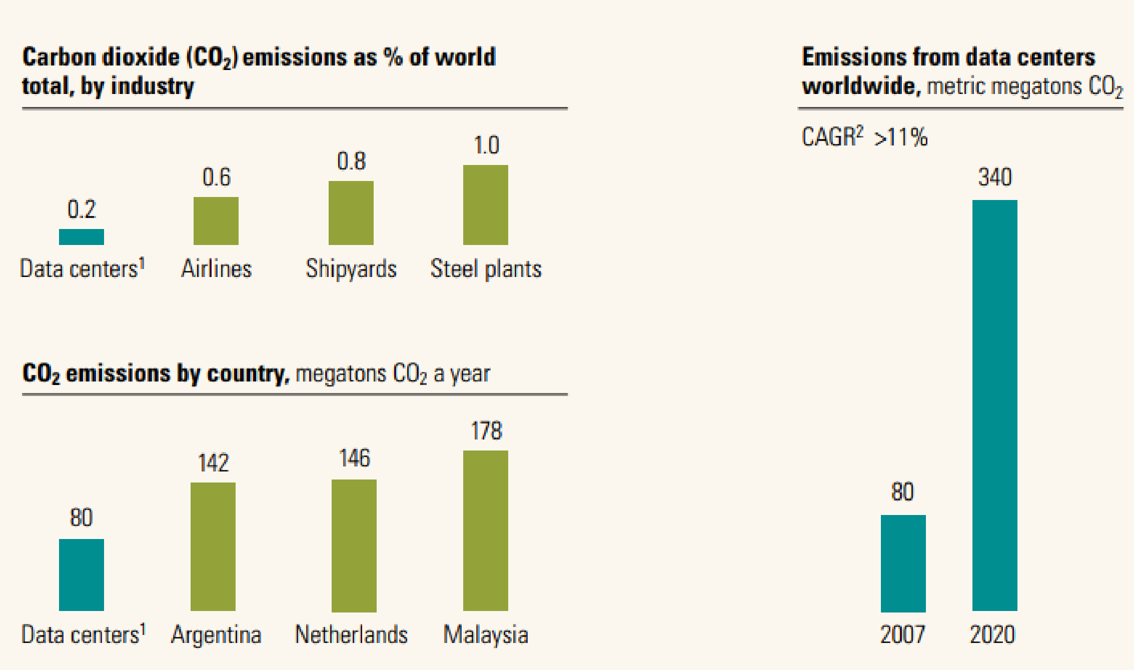


Figure 2.5: Data centers' large carbon footprint [ana07]

One of the important sustainability metrics classifications is the financial aspect, with attention to the financial metrics, [BM10] provides a DC infrastructure and energy cost model in order to compare hardware costs to infrastructure and energy costs. The paper reviews the trends that caused the substantial increase in energy costs and the cost of the DC power and cooling infrastructures. Moreover it provides a methodology for measuring the overall computational efficiency of the DC by describing the two DC efficiency metrics: power usage effectiveness and compute power efficiency and showing the importance of power usage effectiveness with respect to costs.

3 Related Work

While the internet and associated Information and Communications Technologies are diffusing at an astounding pace, DCs proliferate to accommodate this rising demand. But on the other hand, their environmental impacts grow too, therefore much better understanding of DC footprint is urgently needed, in order to reduce the growing environmental impacts. [RMM15] provided a preliminary water footprint accounting for cooling and energy consumption in DCs.

Guidelines are provided in [Bel05] in order to achieve a high level of flexibility and scalability in the DC. These best practices address some issues such as: site location, building selection, and principles in the design and provisioning of critical facilities systems. One guideline related to rack units design, is to use a rack unit as the primary planning factor for estimating the space and power requirements of the DC. Each rack configuration should reflect total power, space and floor-loading demands. Strive for an average of 4 kilowatts per rack across the layout. Another guideline related to planing and designing is to ensure that IT representatives participate in the selection of the architecture and engineering firm that will design and engineer the new DC, and focus particularly on the qualifications, experience and references of the architecture.

For location and site selection, the paper recommended to develop location criteria and use a weighting method to score different locations against high-priority criteria, and to develop a comprehensive site selection checklist that stresses physical security and site flexibility, as well as sets priorities for adequate utilities, public services, and accessibility by employees and service providers, and to build single-story, industrial-type buildings with large floor plates in suburban locations which are the best sites for DCs. Most importantly is the recommendation about power, which recommends to additional capacity into the main electrical components, such as patch panels and conduits, and use higher-gauge electrical wire to accommodate future growth in electrical demand [Bel05]. The following are best practices for power distribution [VSR+17]:

- Assess overall power requirements.
- Strive for multiple utility feeds.
- Provide for maintenance bypass and emergency shutdown.
- Determine if equipment requires single phase or three-phase power.
- Provide for a signal reference grid (SRG) to reduce high-frequency impedance.
- Use power distribution units (PDUs) to integrate circuit breakers and equipment connections.
- Maintain relative humidity levels to minimize electrostatic discharge.
- Be mindful of electromagnetic interference (EMI) and conduct study to determine if shielding or other preventive measures are required.
- Use power-conditioning equipment or integrate into UPS system.

According to [Rom19], The greenhouse gas emissions such as carbon dioxide from fossil fuel-fired electric generators are a major cause of global climate change. The forecast increases in DC electric demand implied a growing source of Greenhouse gas emissions. This is cause for concern. The international community is attempting to reduce Greenhouse gas emissions amid broad scientific and international consensus that climate change is a major threat and must be taken seriously.

Some best practices to achieve a high level of scalability and sustainability in DCs are proposed in [VSR+18]. The paper analyzed the practices of seven DCs in India and the Netherlands and compared the practices followed in these DCs against the relevant standards on sustainable DCs following the dimensions that were presented in [VSR+17], which includes energy efficiency, cooling, air and thermal management, greenness, storage and network. The paper identified design issues and operation inefficiencies, and provided recommendations for improvements in various operation levels of the DCs. Energy efficiency practices enabled them to observe that, decommissioning unused servers results in energy savings of 50%. The best practice for airflow management is to have dedicated horizontal air-flows rather than a mixture of vertical and horizontal air-flows. This paper recommended the DC operators and IT professionals to use infrastructure management or automation tools which can achieve considerable energy savings from 5 to 20%.

An assessment of the environmental impact of DCs was presented in [WASM14]. The paper described the current energy consumption and environmental impact of information technology industry in DCs, and how it is monitored, assessed and benchmarked, and the need for a more holistic approach to manage the environmental impacts in the future. The paper discussed the current manners that approach the environmental issues, and highlighted the need for a change in the currently adopted approaches.

In order to minimize energy consumption in a DC where recirculation of airflow is present, which accordingly leads to inefficiencies in cooling of the racks. An optimization problem is provided in [VDT18], together with characterizing the optimal workload distribution and cooling temperature to achieve minimum energy consumption while ensuring job processing and thermal threshold satisfaction. Moreover, the paper addressed the possibility to uniquely determine the optimal cooling supply temperature and workload distribution as a function of the total workload and desired temperature distribution of the racks in the DC. This was done through analyzing controllers that drive the DC to the optimal state without knowledge of the current total workload to be handled by the DC.

With respect to DCs electricity consumption, an assessment of the growth in DC electricity use from 2005 to 2010 for the US and the world is provided in [Koo+11]. The paper shows that the growth in the installed base of servers in DCs had already begun to slow by early 2007 because of virtualization and other factors, but in those days with the upcoming 2008 financial crisis, the associated economic slowdown, and further improvements in virtualization led to a significant reduction in actual server installed base by 2010 compared to the installed base forecast published in 2007. It shows that electricity used by DCs worldwide increased by about 56% from 2005 to 2010 instead of doubling (as it did from 2000 to 2005), while in the US it increased by about 36%. Moreover electricity used in global DCs in 2010 likely accounted for between 1.1% and 1.5% of total electricity use, respectively. For the US that number was between 1.7 and 2.2%.

One analysis tool that quantifies the overall impact of different energy management approaches during design and operation of DCs, and helps to determine the most cost effective and sustainable approach was provided in [GCS+10]. During design, the tool quantifies the impact of alternatives that can be used to power a DC. During operation, it considers power demand management and evaluates the impact of alternative IT and facilities management policies on time-varying workloads and makes recommendations regarding how much power is required, what mix of power sources is desirable, and how power sources should be allocated across the DC infrastructure.

The growing need to eliminate the DC environmental issues on the planet lead to the assessment methods that include DC metrics, which focus on single issues during their operation; and building environmental assessment methods, which consider multiple issues for the whole facility. A screening Life Cycle Assessment is applied in [MMS+12], the paper describes a methodology that enables engineers at early stages of the design process to quantify multiple environmental impacts, throughout the full supply chain of a DC. The Authors of this paper used information from an existing DC as a vehicle for their study. The paper shows that the results of their life cycle assessment can not only be used to improve the environmental performance of a DC, they can also help in reducing financial and operational expenditure, such as eliminating the need for mechanical cooling can reduce the need for certain pieces of equipment; the production and running of which impacts on the environment, and can in turn reduce overall costs. Generally, the impacts of decisions on capital and operational expenditure can be a good motivation for reducing environmental impact, but cost implications are not always an evident and should not be considered in isolation from environmental concerns.

Current DC simulators focus more on the resource management and deployment of applications. For example, DCSim provides an application model that can simulate the interactions and dependencies between many Virtual machines working together. It supports other features of virtualization, such as a work conserving Central Processing Unit (CPU) scheduler used in modern hypervisors, resource allocation, and Virtual Machine (VM) migration and replication. Moreover, it allows host power states (on, off, suspended) to be modelled with appropriate transition times between states. DCSim was designed specifically to study VM management in a DC providing an Infrastructure as a service (IaaS) Cloud [TKBL12]. GDCSim is another tool that testes different DC geometries, workload characteristics, platform power management schemes, scheduling algorithms and DC configurations to simulate a DC, this tool supports automated processing that allows interfacing different modules together and does not require user intervention once the simulation has started [GGB+11].

Generally, DCSim supports Virtualisation modeling which is not supported by many of the other simulators. GDCSim focuses on scheduling algorithms, DC configurations and automated processing. The other aforementioned researches in [VSR+18], [Bel05], provide best practices and guidelines for resolving DC issues without employing such guidelines on a simulator to assess the feasibility. The screening Life Cycle Assessment in [MMS+12] describes a methodology that enables engineers at early stages of the design process to quantify multiple environmental impacts, but it does not provide a simulation for this methodology.

But most of the existing simulators use a large number of inputs to evaluate the sustainability of a DC. Therefore, the goal of this research is to model a DC simulator, which can consider a limited number of commonly reused inputs and evaluate the maximum number of sustainability metrics to measure the different key performance indicators, which consequently can lead to more efficient and sustainable DCs in the future. A sustainable DC simulator is created, which focuses on achieving sustainability in DCs based on using the minimum number of inputs that are selected precisely according to analysing the frequency of usage for each input.

4 Methodology

The research design of this thesis is the casual design, because the core of this work, is to apply an explanatory study [BP10], to analyse the interrelationships between the input parameters and metrics of a DC to identify the minimum number of commonly reused inputs that can be used to evaluate the maximum number of sustainability metrics, which are classified in [VSR+17]. This thesis aims at explaining how changes in the workload will impact the sustainability of a DC.

This work seeks to accomplish a deep analysis of the DC inputs and metrics. Tree graphs are used to demonstrate the connection between the inputs and how they are used to evaluate the metrics. Afterwards, a quantitative research is applied to identify the commonly reused inputs and discard the other inputs based on the frequency of repetition of each input. Neatly, each commonly reused input is then represented in a rooted tree graph [GY05], in which the root node is the commonly reused input, and it has a set of edges going into siblings, which are the the metrics that use this input to be evaluated. Then, we identified the minimum number of commonly reused inputs that can evaluate the maximum number of sustainability metrics.Afterwards, we compared the existing simulation platforms, to check if there is any simulator that can simulate a DC with this minimum set of commonly reused inputs, but the comparison shows that no simulator fulfill the requirements. Consequently, we designed the customized sustainable DC simulator to receive the input values of the commonly reused inputs, and evaluate the sustainability metrics according to the incorporated equations.

The evaluation of the maximum number of sustainability metrics has to reflect the sustainability of the DC and help DC operators to monitor the change in workload and how it affects the DC sustainability. It was impossible to find a realistic sample of DC inputs to use it in the experiments because DC operators don't publish data about their DCs. Therefore, a hypothetical sample of DC inputs was used in the experiments. Finally, a discussion of the analysis and simulation results is provided, along with the results visualization.

5 Sustainability Metrics Analysis and Results

The goal of this research is to model a DC simulator, which can make use of the minimum number of inputs and evaluate the maximum number sustainability metrics to measure the different key performance indicators, which consequently can lead to more efficient and sustainable DCs in the future. Therefore, this thesis is intended to apply a thorough analysis on the available sustainability metrics, and research the inputs that could help in estimating the the metrics that give insights about the environmental impacts and the power consumption level of a DC. Understanding how changes in the input values of a DC are reflected in the metrics, requires analysis of metrics inputs of a DC.

5.1 Analysis of Metrics and Inputs

A table is provided in Appendix A, which lists all the input names, measurement units and a description of each input to not to repeat any of the inputs [VSR+17]. The first step in the analysis of the sustainability metrics is to form a directed tree graph for each category of metrics according to the taxonomy of metrics proposed in [VSR+17]. In general, tree graph consists of a set of objects which are called vertices or nodes that are connected together by edges, where all the edges are directed from one vertex to another [BM+76]. We are using the same idea of nodes and edges to model the DC inputs and metrics. Therefore, tree graph is adopted to be used broadly in the analysis.

Each tree graph consists of a set of inputs represented by a rectangle and a set of metrics represented by an oval shape. The edges are used to connect the inputs and the corresponding metrics, the arrow of the edge shows which input is used to evaluate which metric. The second step is to determine the subset of inputs and metrics that are commonly reused, if some inputs or metrics are commonly reused, this means that these inputs or metrics have more effect on the metrics values, in other words the commonly reused inputs values are used broadly to evaluate the key indicators of a DCs, and if there is a change in these inputs, it leads directly to change in the sustainability metrics. Hence, these commonly reused inputs and metrics are used to model a sustainable DC simulator in this research.

In order to highlight the commonly reused inputs and metrics, a coloring scheme is used, the commonly reused inputs are highlighted in yellow, which means that if we see a yellow rectangle, then it represents an input that has been reused more than once to evaluate another input or metric. Similarly, the commonly reused metrics are highlighted in blue, which means that this blue oval shape represents a metric which is reused more than once to calculate another input or metric.

Figure 5.1 shows the graph notation symbols that are used for creating the tree graphs:

- White rectangle: represents a normal input.
- Yellow rectangle: represents a commonly reused input.
- White oval: represents a normal metric.
- Blue oval: represents a commonly reused metric.
- Arrow: connects inputs and metrics.

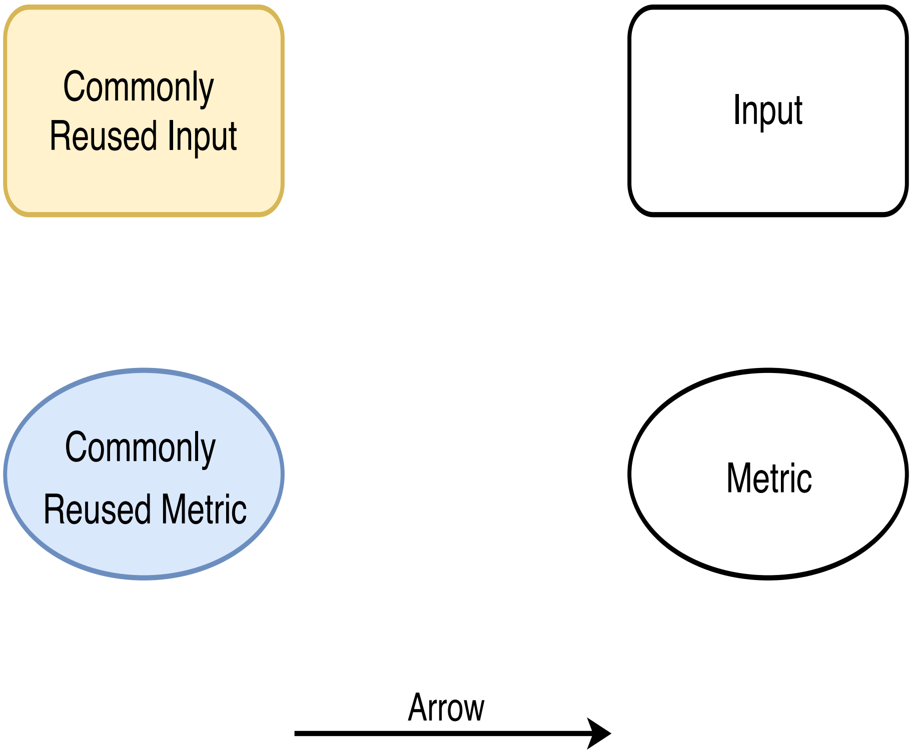


Figure 5.1: Graph notation symbols

The next step after listing the set of inputs and providing the graph notation, is to create a graph for each category of metrics according to the taxonomy of metrics proposed in [VSR+17], these tree graphs show how the inputs are used for calculating the different inputs and metrics. The input which is represented by a white rectangle is connected to an oval node which represents a metric, and an arrow is used for connecting the inputs and the metrics. A tree graph is created for each of the following metrics categories according to the taxonomy of sustainability metrics proposed in [VSR+17]:

- Energy efficiency metrics.
- Cooling metrics.
- Greenness metrics.
- Performance and productivity metrics.
- Thermal and air management metrics.
- Network metrics.
- Storage metrics.
- Security metrics.
- Financial metrics.

Energy Efficiency Metrics

Electricity generation is one of the three principal sources of greenhouse gas emissions, while transportation and deforestation are the other two [Bou11]. In order to reduce a DC's electrical consumption and as a consequence reduce greenhouse gas emissions, it is hard to link DC activities to electricity use, as a DC consumes electricity when IT and physical infrastructure loads are using power to process information [Bou11].

When considering energy efficiency metrics, it is very important to differentiate between power and energy, a brief explanation about both terms is given, energy is measured by joules or watt per second and it is defined as the capacity to do some activities or work such as moving a car or operating a machine, on the other hand power is measured by kilowatt (kW) and it is defined as the rate of generating energy, it can not be converted from one form to another while energy could be transformed from one form to another.

Since there are a large number of energy inputs and metrics, we split the graph into two parts. Figure 5.2 shows the graph of the first part of energy efficiency metrics, the yellow rectangles represent the inputs which are commonly reused more than once to compute a specific input or metric, while the blue ovals represent the metrics that are commonly reused more than once in order to compute a specific input or metric, this coloring scheme is adopted in all the graphs to easily recognise the commonly repeated inputs and metrics.

The input "Total Facility Energy" is repeated 3 times only in Figure 5.2, it is together with the input "Total Hardware Compute Load DC" to evaluate the metric "IT Hardware Power Overhead Multiplier (H-POM)". Additionally, it is reused with the input "EDC Useless Work" to evaluate the metric “Energy Wasted Ratio (EWR)”. Therefore, we consider the input “Total Facility Energy” to be a commonly reused input. Figure 5.3 shows that the metric “PUE” is colored blue since it reused three times to evaluate other metrics. Therefore, we consider it to be a commonly reused metric. The list of commonly reused inputs in Figure 5.2 are:

- ITEU.
- Total facility energy.
- IT energy.
- Total deployed servers.
- Planned Energy (EPi).
- Total facility power.
- IT power.
- Total DC space.
- Performance.

While the list of commonly reused metrics in Figure 5.2 are:

- PUE.
- IT Energy Efficiency (ITEE).
- IT Usage Effectiveness (ITUE).

Representing all inputs and metrics for each category together in one graph, helps in quickly finding the commonly reused set of inputs and metrics just by looking on the graph. Therefore, we can focus on these inputs and metrics while modeling the DC simulator, and neglect the others that colored in white. Moreover, it shows how each metric is evaluated from it’s connected nodes which represent the inputs.

PUE and Data Center Infrastructure Efficiency (DCiE) are two major sustainability metrics for evaluating efficiency in the DC. PUE is a ratio of the total power coming into the DC to the power going to the IT. For example, if it takes 1.5 times as much energy to run your DC than is required for just the IT equipment, the PUE would be 1.5. Smaller PUE values indicate a more efficient DC. DCiE is represented as a percentage of IT load power divided by total DC power, and is the inverse of PUE or 1/PUE [Elec].

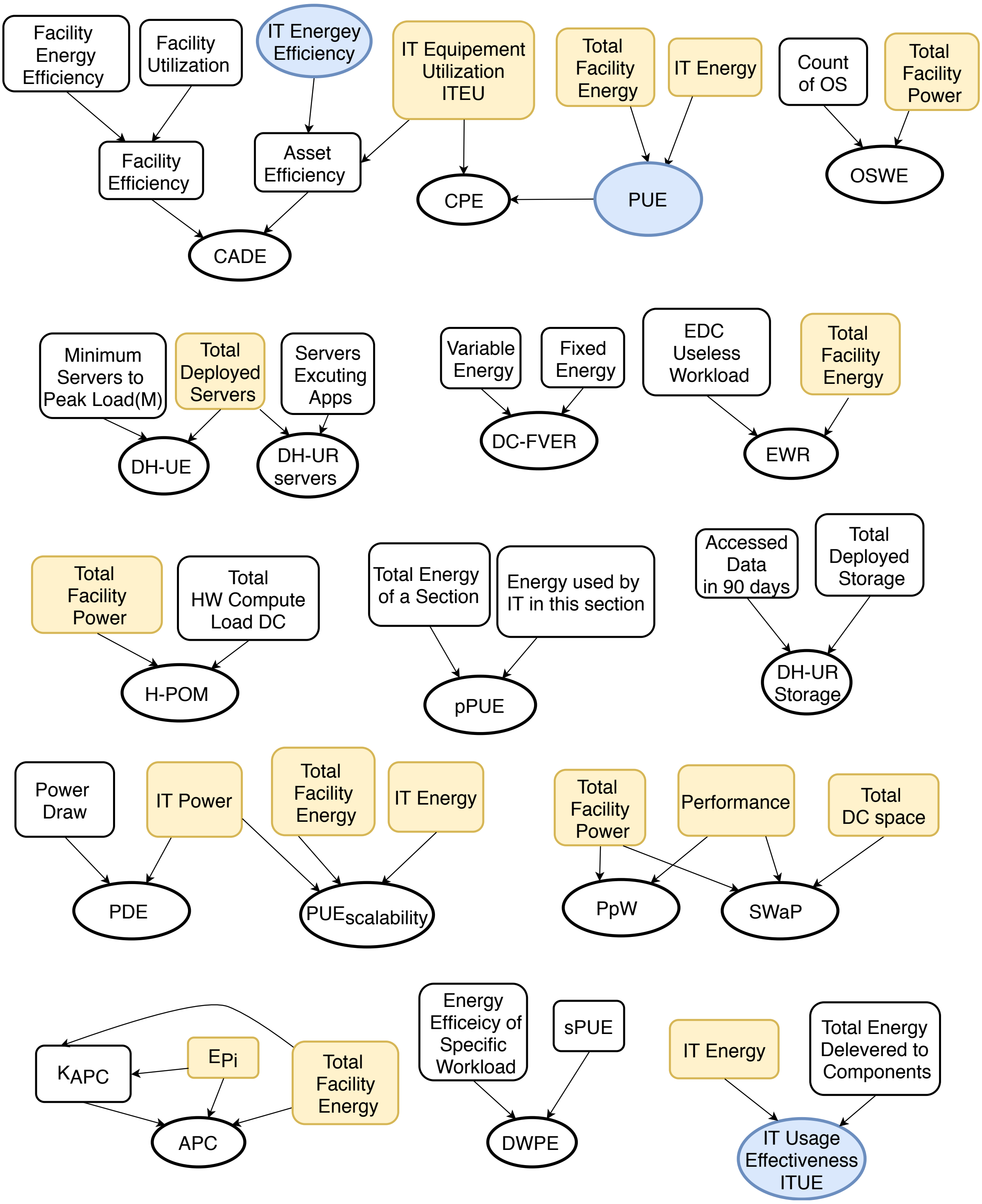


Figure 5.2: Energy efficiency metrics part I

Figure 5.3 shows the rest of energy efficiency metrics, it is clear that the following set of inputs are commonly reused :

- EDC Real i.
- EDC Baseline i.
- E oth DCi base.
- EDC i current.
- Cost ei current.
- Useful work.

Likewise, Figure 5.3 shows that "Green Energy Coefficient (GEC)" and "ScE" are commonly reused metrics:

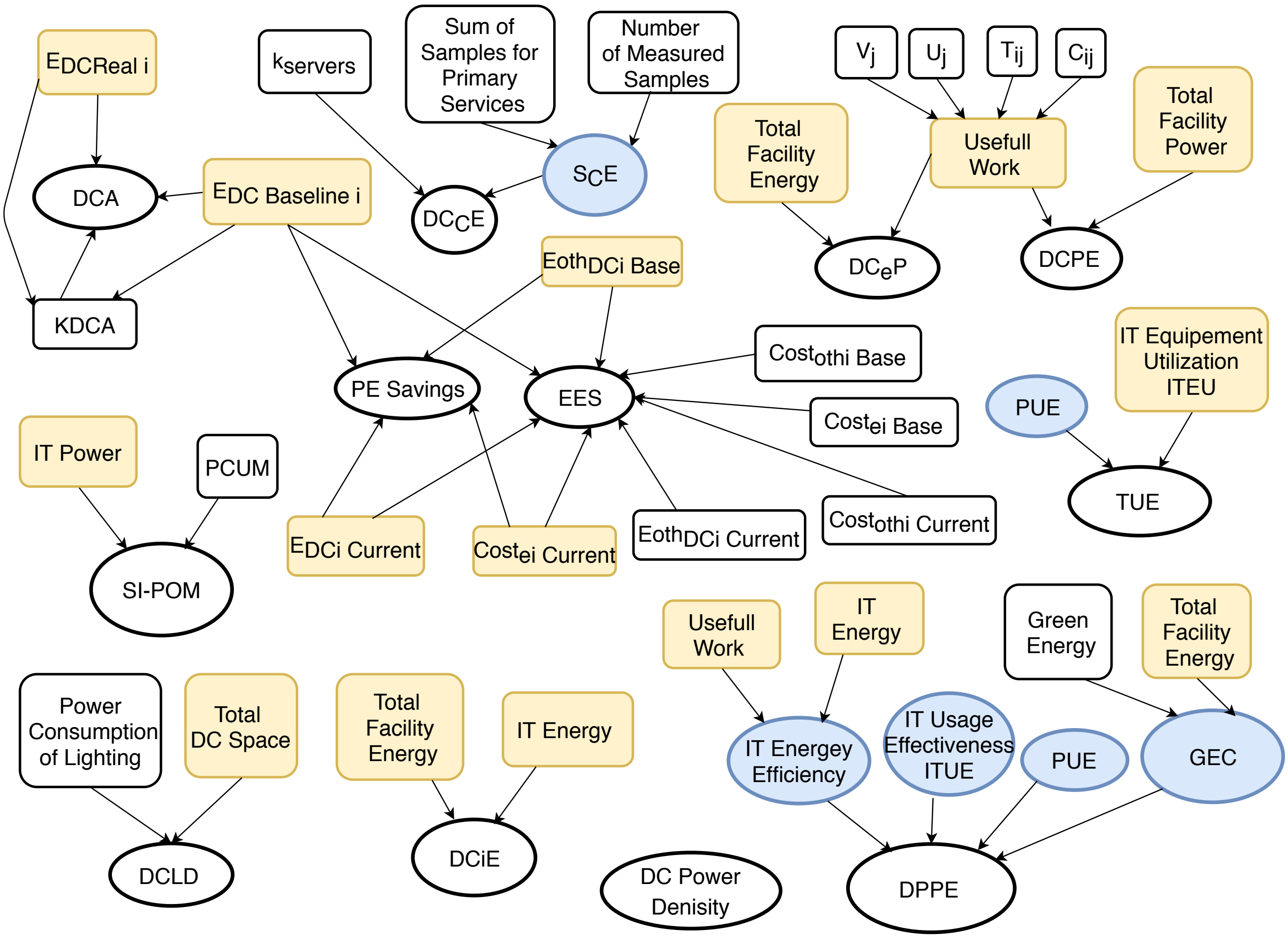


Figure 5.3: Energy efficiency metrics part II

Cooling Metrics

Figure 5.4 graph shows the cooling metrics such as DC cooling system efficiency, DC cooling system sizing factor, energy efficiency ratio, heating, ventilation, and air conditioning (HVAC) system effectiveness, and air economizer utilization factor, and how they are evaluated using the different inputs. The input “IT Energy” is the only commonly reused input.

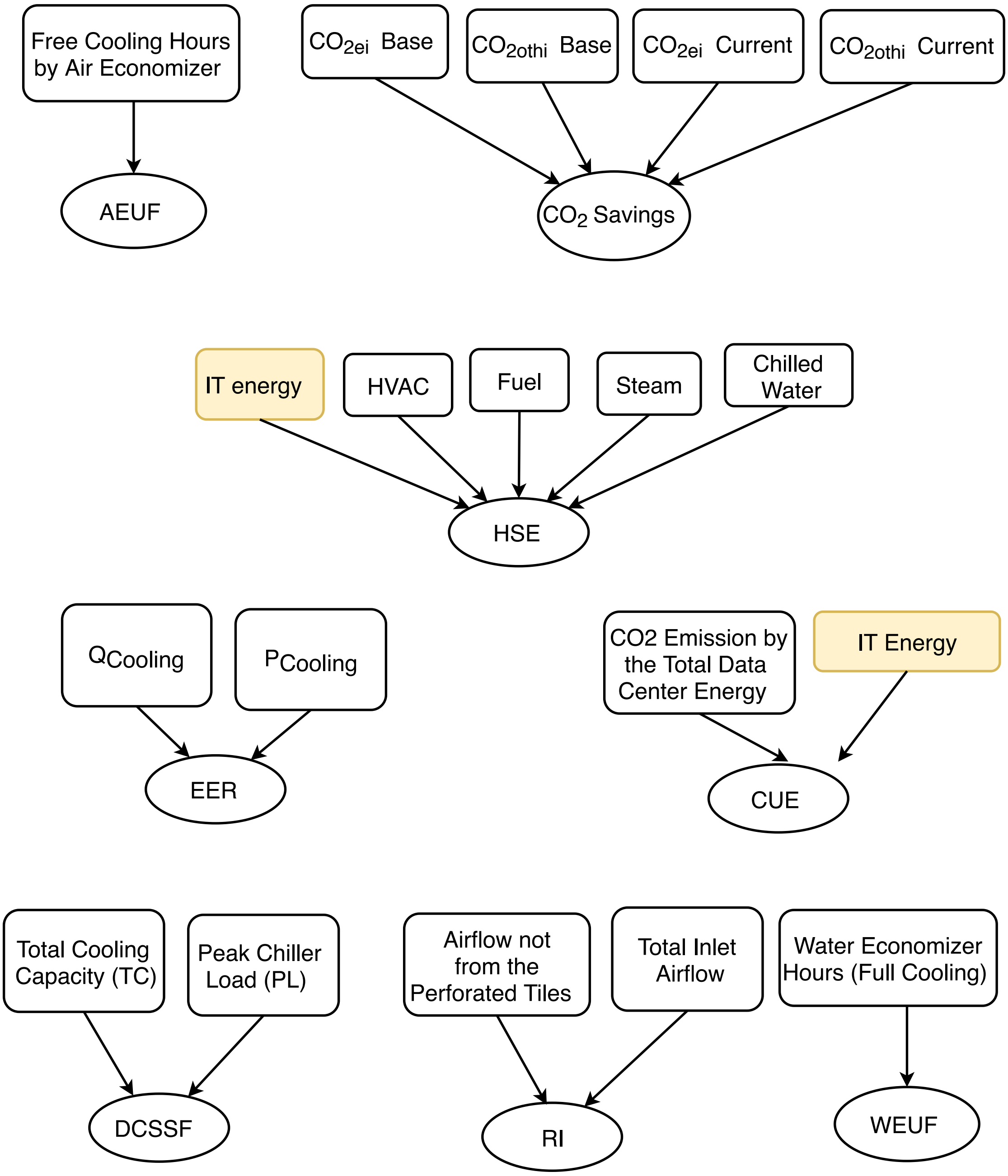


Figure 5.4: Cooling metrics

Greenness Metrics

A green DC is used to achieve maximum energy efficiency and minimum environmental impact [Mur08]. Greenness metrics such as carbon usage effectiveness, energy reuse effectiveness, green energy coefficient, CO2 savings, electronics disposal efficiency, energy reuse factor, material recycling ratio, technology carbon efficiency, water usage effectiveness, the green index, and others are all collectively intended to evaluate the greenness of the DC in terms of carbon footprint, heat reuse, efficiency of water consumption and use of renewable energy resources.

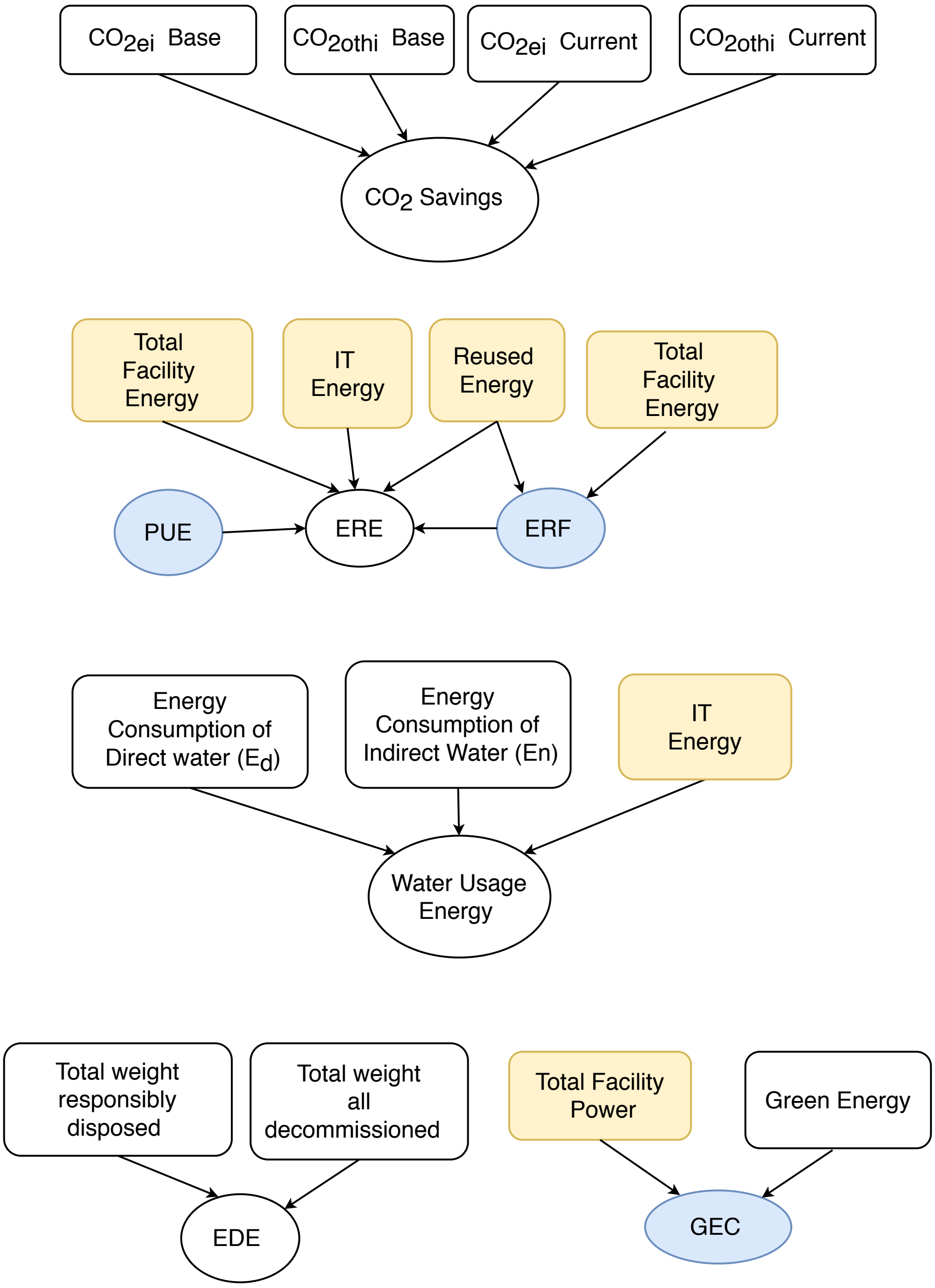


Figure 5.5: Greenness metrics part I

Figure 5.5 shows the first part of inputs and metrics in Greenness category, and it shows that the following set of inputs are commonly reused:

- IT energy.
- Total facility energy.
- Total facility power.
- reused energy.

Similarly, Figure 5.5 shows that the following set of metrics are commonly reused:

- PUE.
- Energy Reuse Factor (ERF).
- GEC.

Figure 5.6 shows the second part of inputs and metrics in Greenness category, and it indicates that the following set of inputs are commonly reused, and there are no commonly reused metrics:

- IT energy.
- IT power.
- Total facility power.

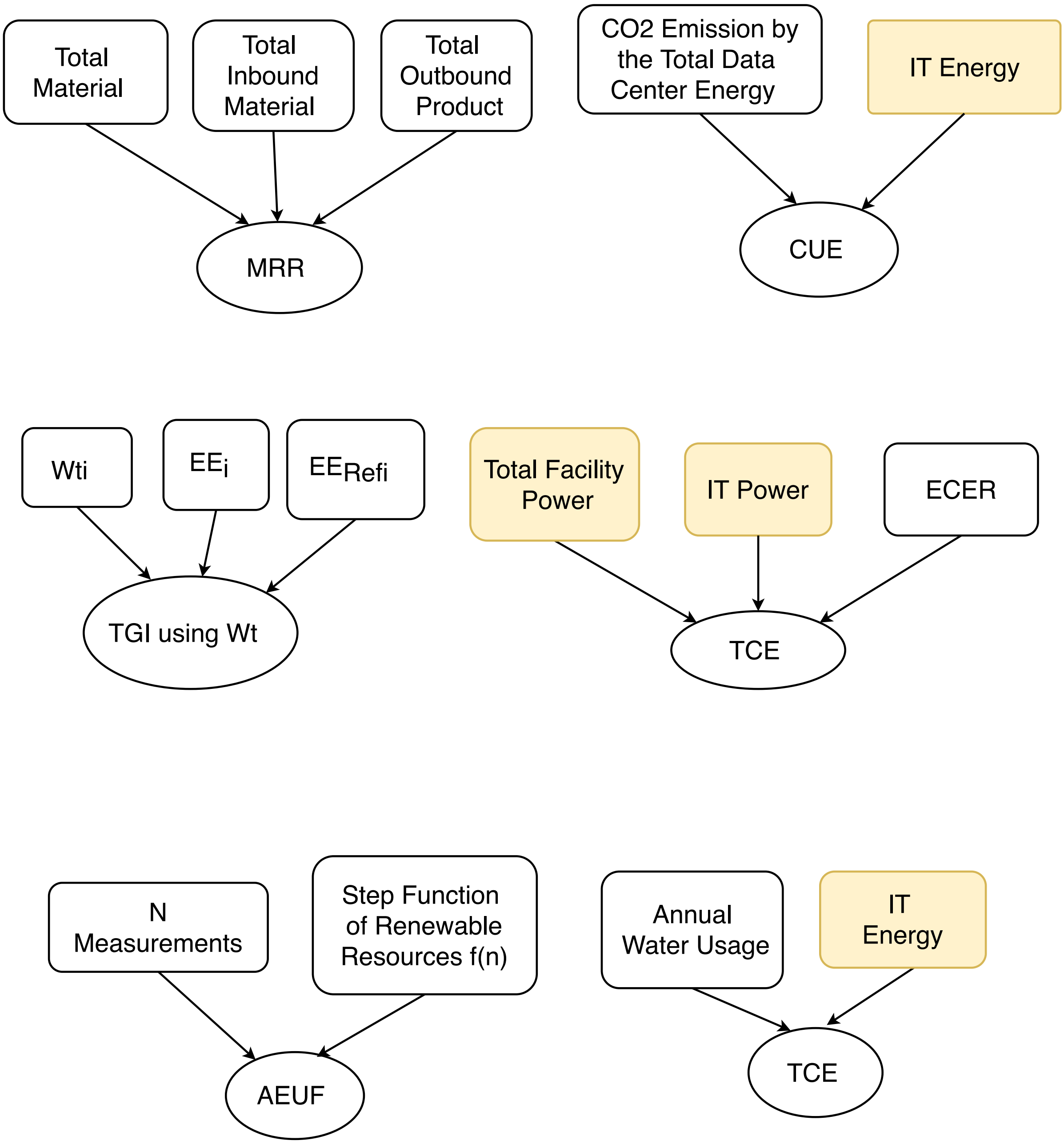


Figure 5.6: Greenness metrics part II

Performance and Productivity Metrics

Performance and productivity metrics such as Central Processing Unit Usage, DC energy efficiency and productivity index, energy proportionality, energy proportionality, availability, capacity, and efficiency performance score, DC productivity, server and DC utilization, and others are important in terms of determining the mostly significant workloads and energy distribution losses, and measuring the energy consumption at various operational modes of the DC. These metrics help DC operators to measure and evaluate the performance and productivity, which in turn assist the DC operators in the planning and expansion process [VSR+17].

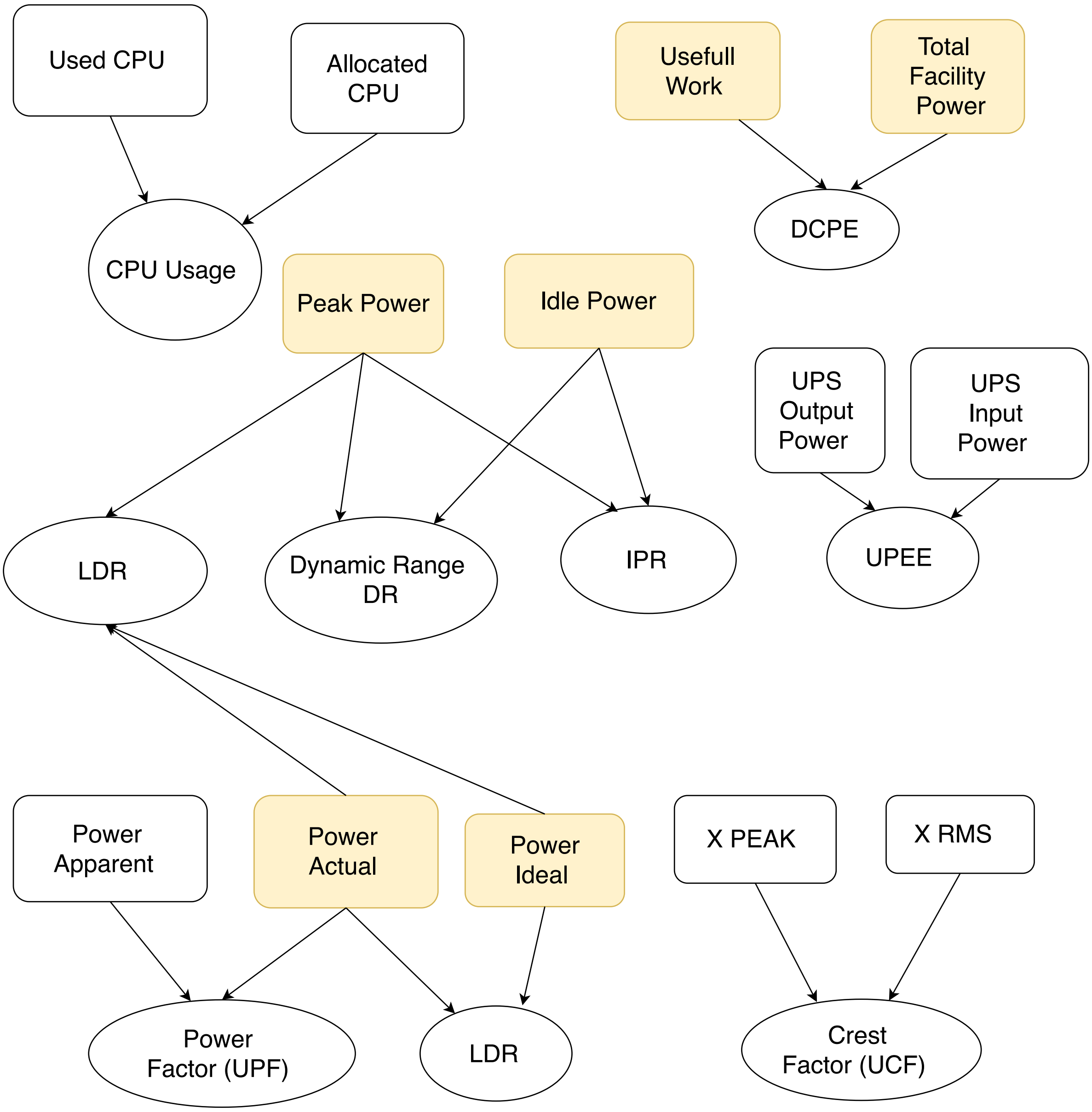


Figure 5.7: Performance and productivity metrics part I

Figure 5.7 shows the first part of inputs and metrics in performance and productivity category. It shows that the following set of inputs are commonly reused, and no reused metrics:

- Useful work.
- Idle power.
- Peak power.
- Ideal Power.
- Actual Power.

Similarly, Figure 5.8 shows that the following set of inputs are commonly reused and no common metrics:

- Total facility power.
- IT power.

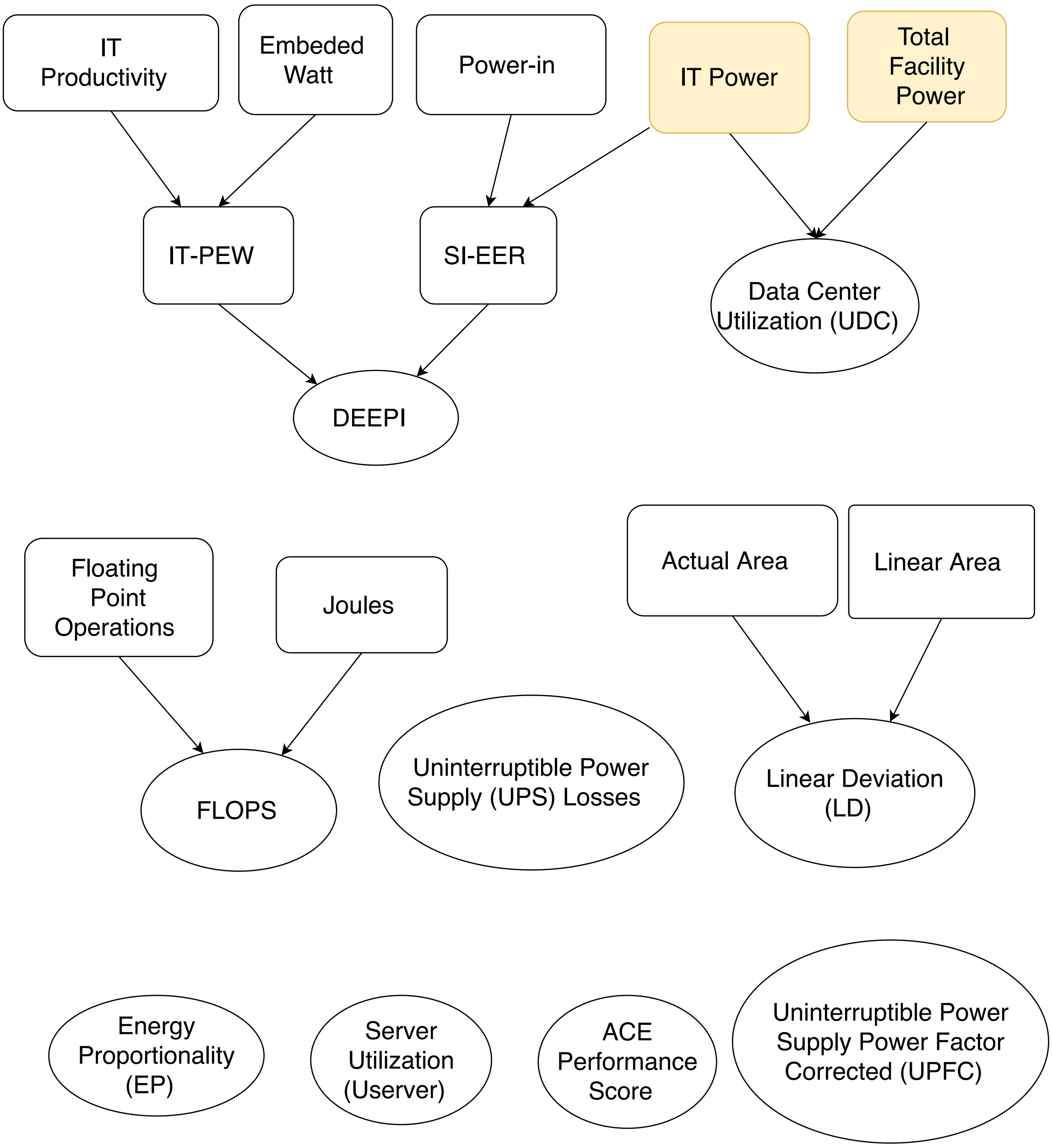


Figure 5.8: Performance and productivity metrics part II

Thermal and Air Management Metrics

Thermal and air management metrics such as DC temperature, airflow efficiency, balance ratio, bypass ratio, heat flux, imbalance of temperature, rack cooling Index, relative humidity, recirculation Ratio, and others are intended to measure environmental conditions of the DC and also determine how air flows within a DC, from cooling units to the vents [VSR+17]. Figure 5.9 shows that the input "Server Airflow(Ms)" is a commonly reused input, while the metric "Computer Room Air Conditioning (CRAC)" is a commonly reused metric.

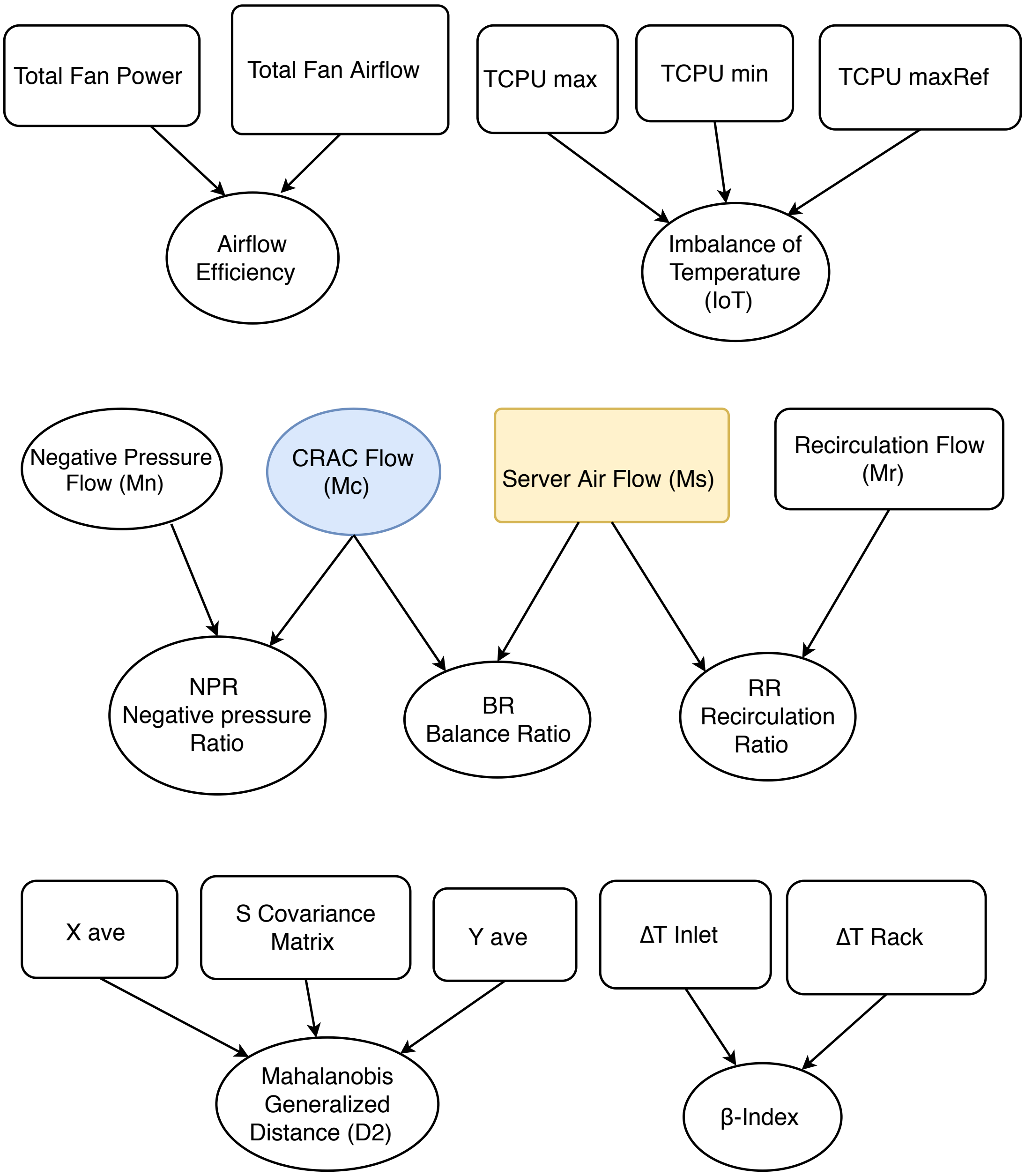


Figure 5.9: Thermal and air management metrics part I

Figure 5.10 indicates that is no commonly reused inputs or metrics.

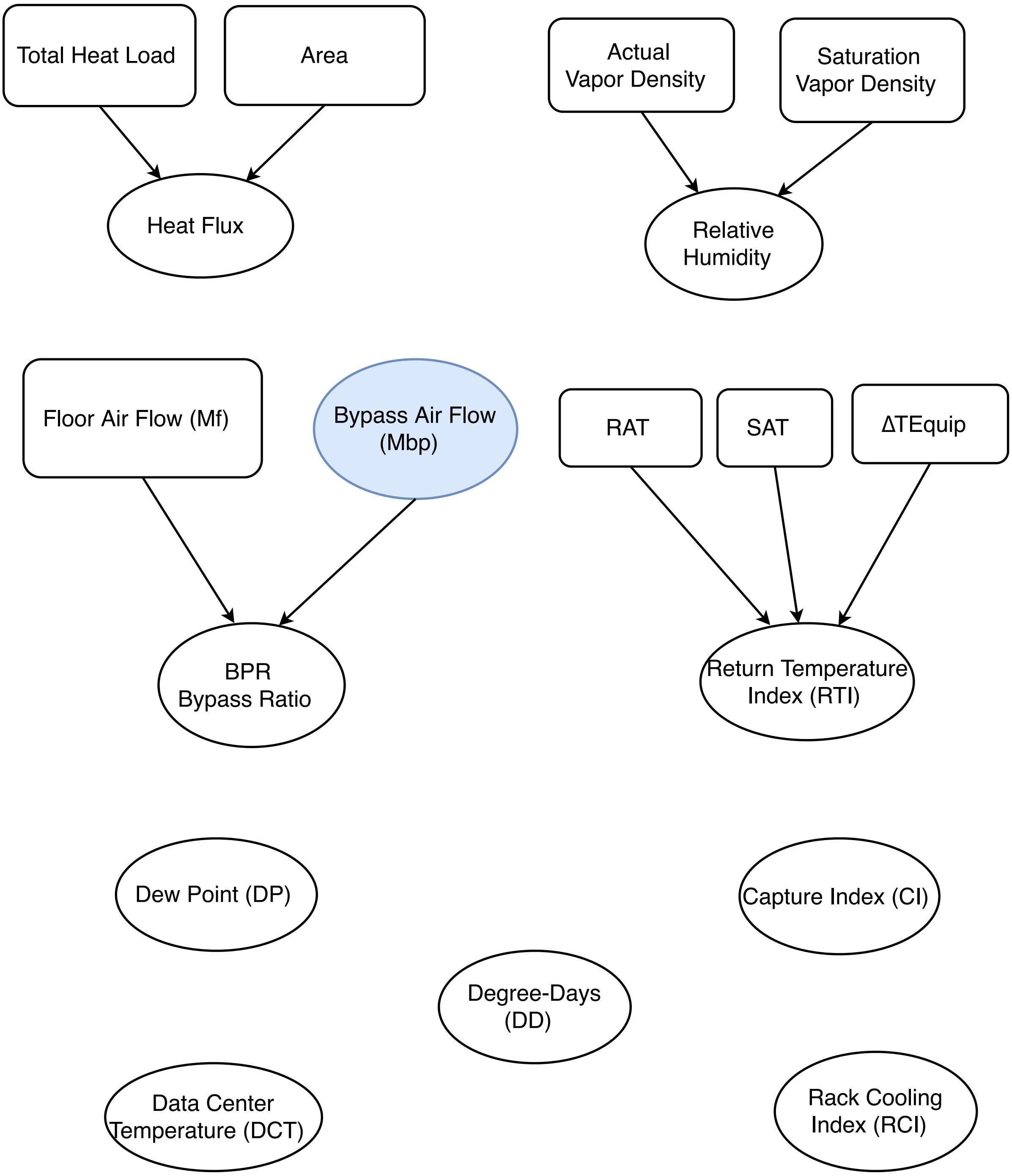


Figure 5.10: Thermal and air management metrics part II

Network Metrics

Network metrics such as communication network energy efficiency, diameter stretch, energy consumption rating variable load, network power usage effectiveness, and others can be used to evaluate the network efficiency of a DC. Figure 5.11 shows that the inputs “Useful Work”, “Total Energy Facility” and “IT Power” are commonly reused, and there are no commonly reused metrics.

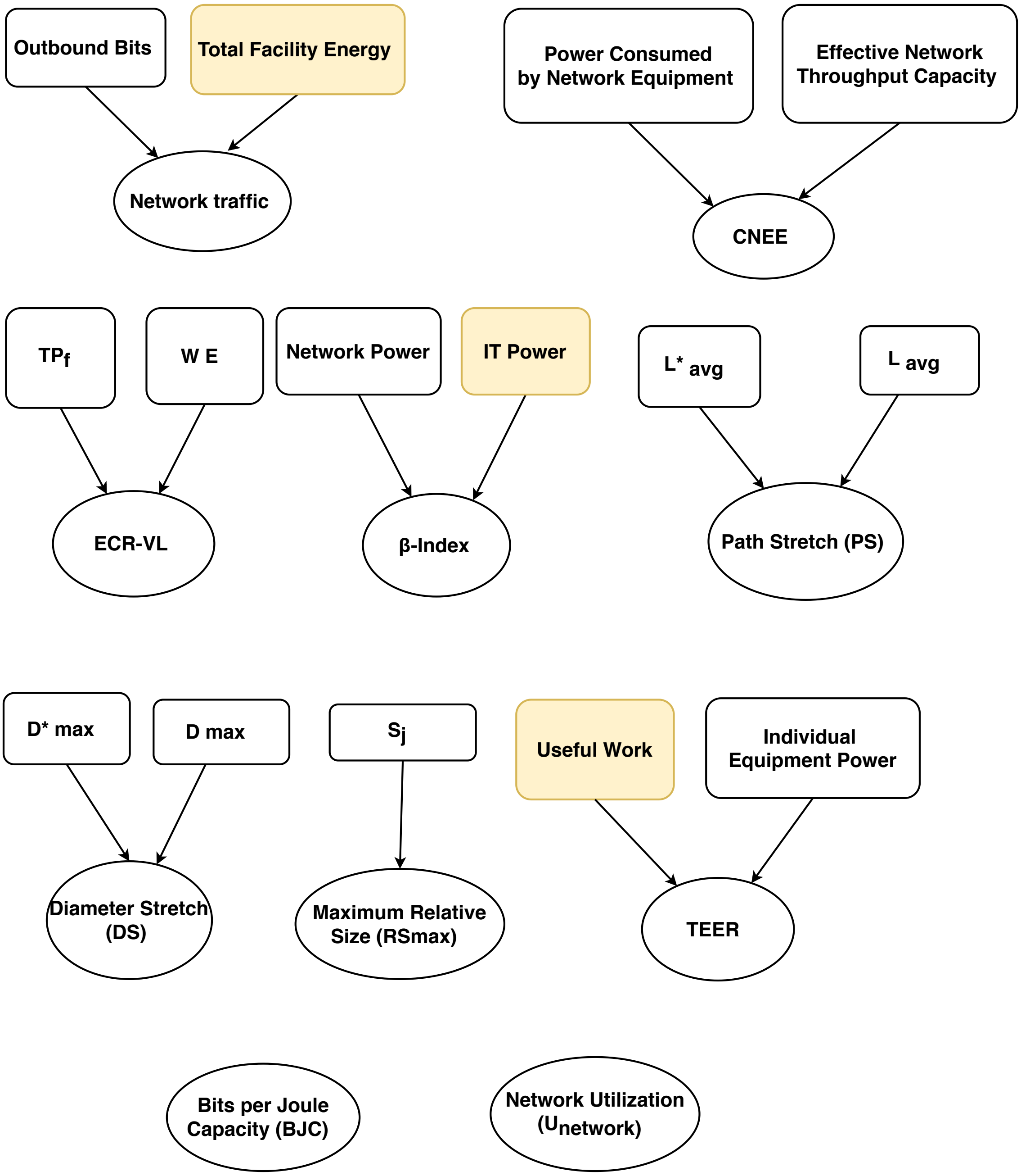


Figure 5.11: Network metrics

Storage Metrics

Storage metrics such as storage usage, slot utilization, overall storage efficiency, low-cost storage percentage, capacity, memory usage, response time, and throughput can be used for storage utilization and the suitable capacity allocation. Figure 5.12 shows that the inputs "Total Raw Memory Capacity" and "Used Memory by a Server or Application" are commonly reused, and there are no commonly reused metrics.

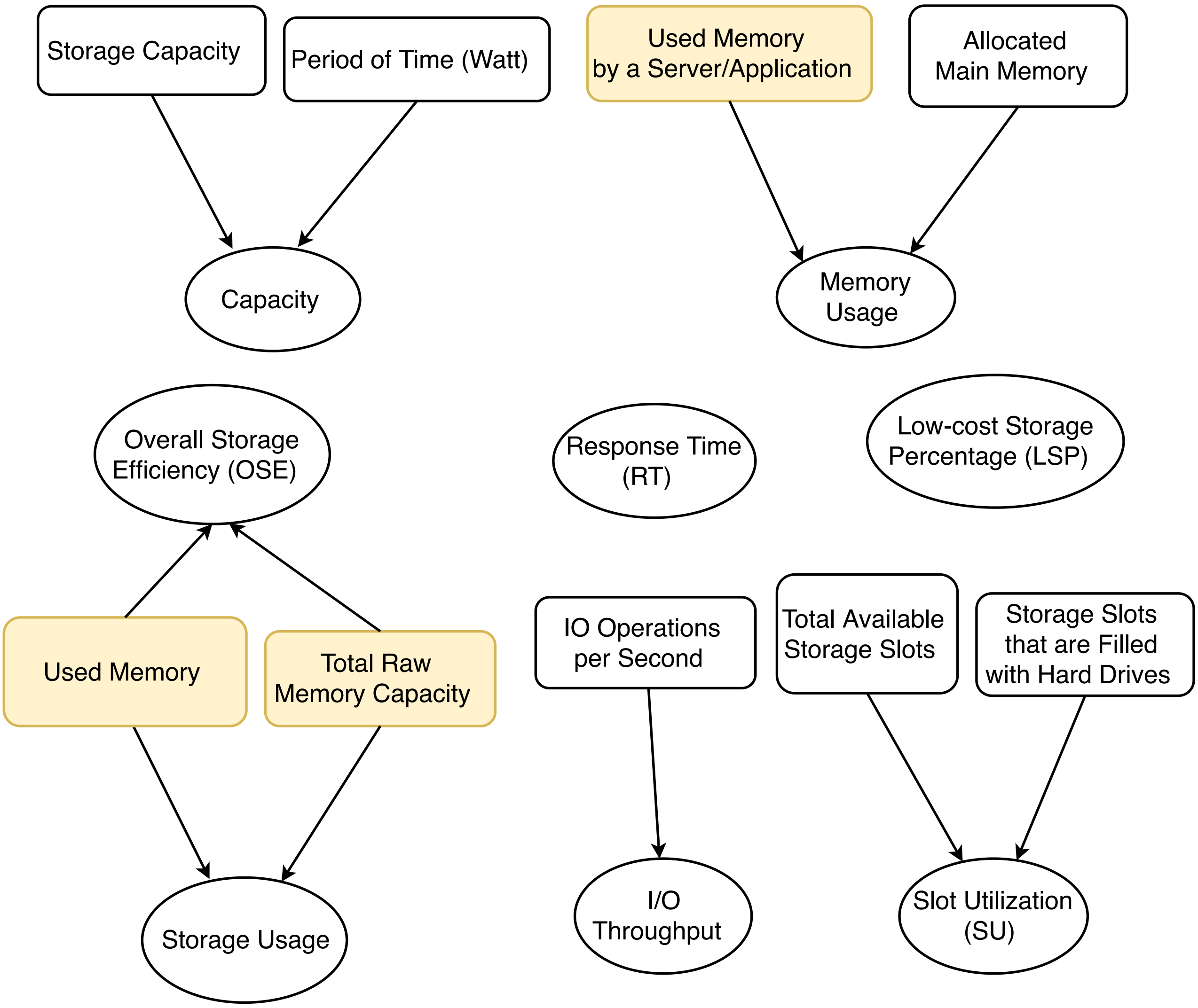


Figure 5.12: Storage metrics

Security Metrics

Security metrics such as average comparisons per rule , accessibility surface, application transaction rate, concurrent connections, connection establishment rate , connection tear down rate, defense depth, detection performance, data transmission exposure, firewall complexity, transfer rate, interface accessibility surface, fragmentation handling, and others help in assuring the level of authentication, authorization and data protection in a DC. Figure 5.13 shows that there are no commonly reused inputs or metrics.

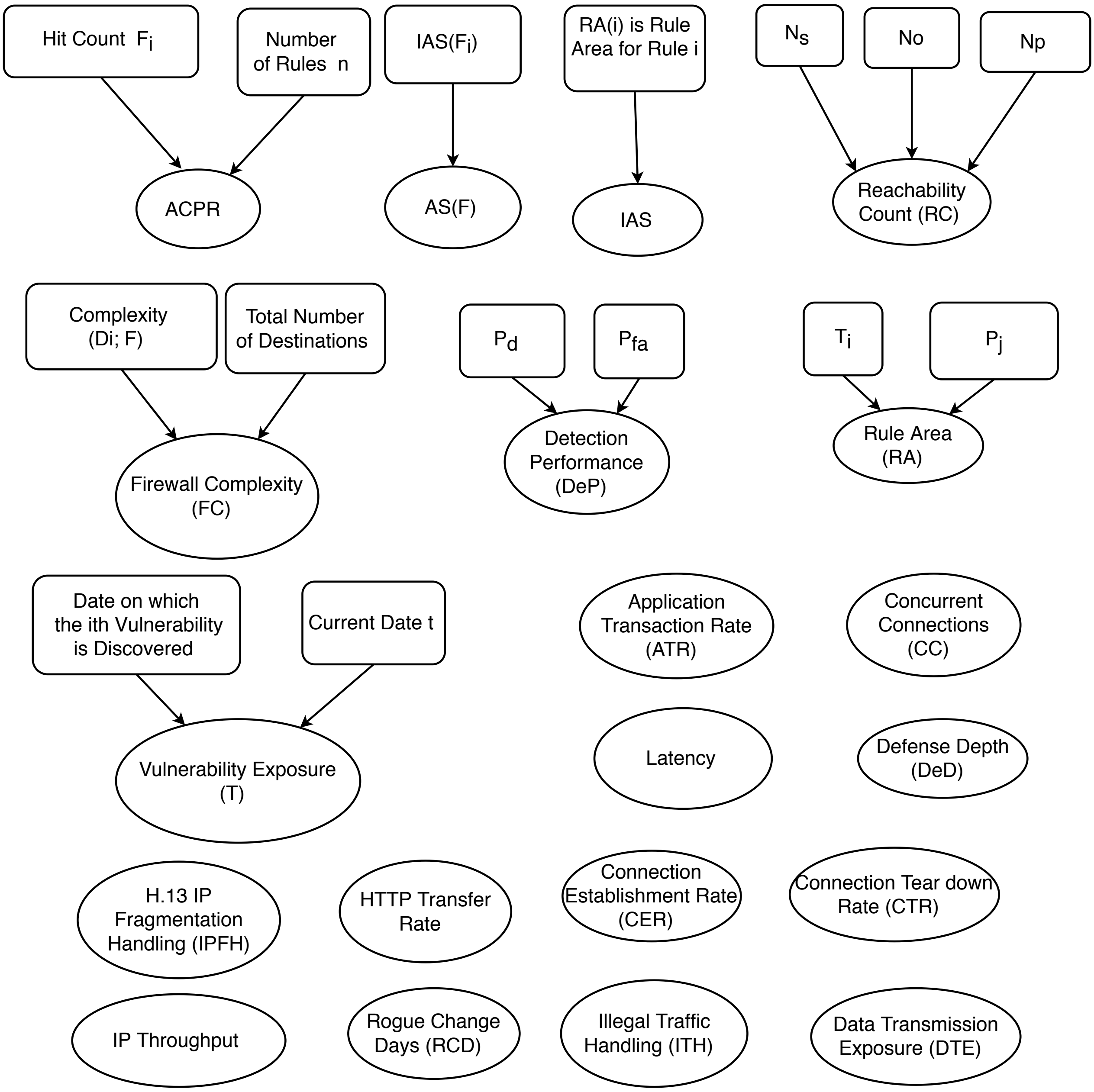


Figure 5.13: Security metrics

Financial Metrics

Financial metrics such as business value of converged infrastructure, capital expenditure, total cost of ownership, operational expenditure, return on investment, availability, and others are used for calculating total cost of ownership, financial impact of DC outages, return on investments on management tools and technologies for sustainable DC [VSR+17]. Figure 5.14 shows that there are no commonly reused metrics or inputs.

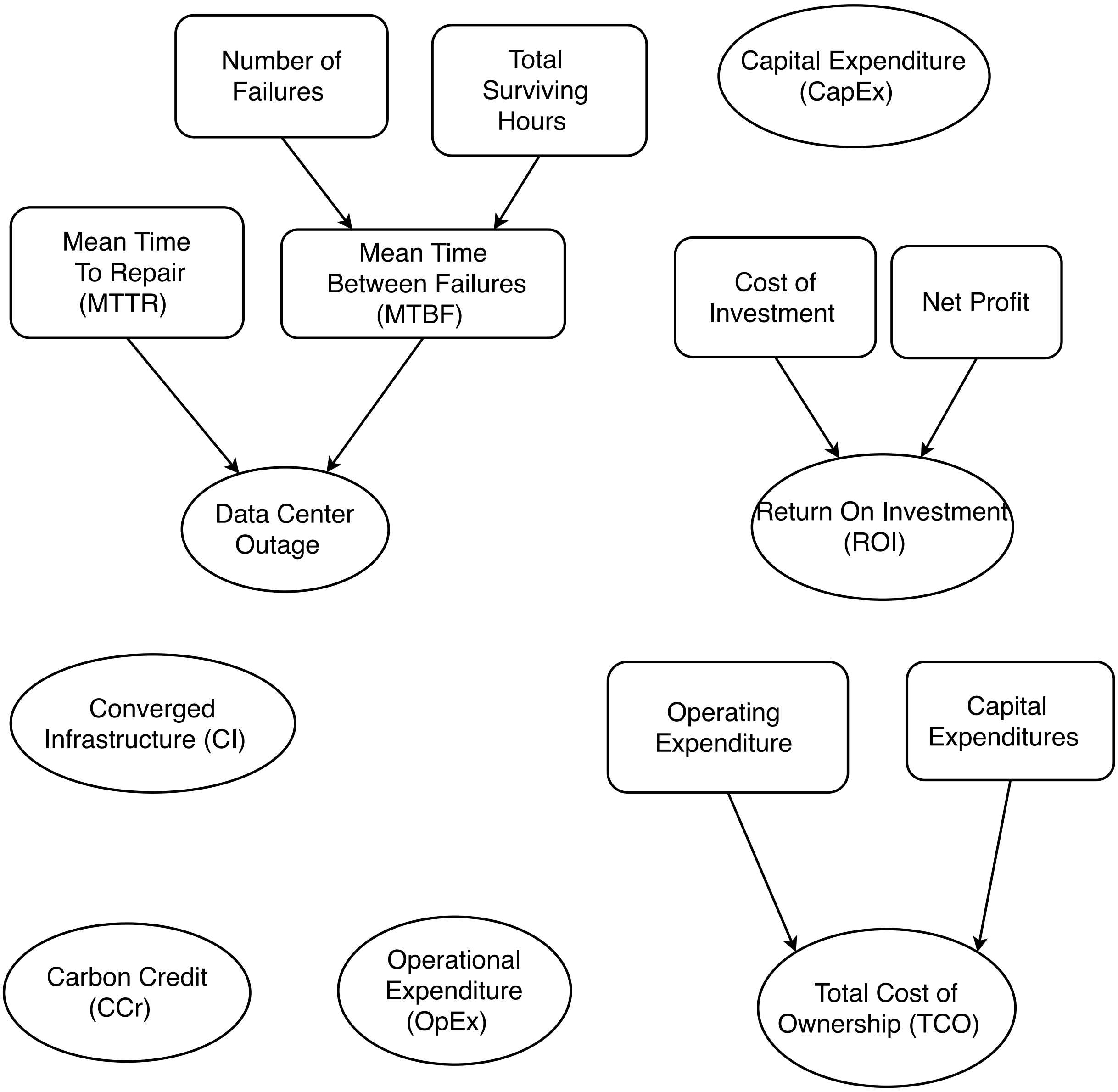


Figure 5.14: Financial metrics

Commonly Reused Metrics and Inputs

Once the graphs for each category of the sustainability metrics are created, a summary of the commonly reused inputs and metrics can be provided. Table 5.1 summarizes all the commonly reused inputs, it includes each input and it's frequency. The frequency in this table represents the number of edges that are related to this inputs, it indicates how many times is this input reused to calculate other inputs or metrics.

Table 5.1: The list of commonly reused inputs and their frequencies.

Input Name	Frequency
IT Energy	14
Total Facility Energy	12
Total Facility Power	7
IT Power	7
Useful Work	6
EDC Baseline i	4
Actual Power	3
Peak Power	3
Used Memory by Servers or Application	3
Planned Energy EPi	2
IT Equipment Utilization ITEU	2
Total Deployed Servers	2
Total DC space	2
Performance	2
Cost ei current	2
E othDCi base	2
EDC i current	2
EDC Real i	2
Idle power	2
Reused Energy	2
Ideal Power	2
Server Air Flow	2
Total Raw Memory Capacity	2
Total Energy Delivered to Compute Components	2

Similarly, Table 5.2 summarises the set of commonly reused metrics and their frequencies.

Table 5.2: The list of reused metrics and their frequencies.

Metric Name	Frequency
Power Usage Effectiveness (PUE)	4
IT Energy Efficiency	2
CRAC Flow (Mc)	2
IT Usage Effectiveness (ITUE)	1
Server Compute Efficiency (SCE)	1
Green Energy Coefficient (GEC)	1
Energy Reuse Factor (ERF)	1

From Table 5.2, it is known for example, that the input "Total Facility Energy" is used 13 times as input for other inputs or metrics, while the input "Total Raw Memory Capacity" is repeated twice , and from Table 5.3, it is known that the metric "UPS" is used 4 times as input for other metrics or inputs, while the metric "Negative Pressure Flow" is repeated only one time.

The reasoning and scientific foundation around the exclusive selection of these inputs and metrics can be explained using graph theory. Therefore, a graph that consists of nodes representing the inputs or metrics, and their relationships as edges will be formed for each input and metric in Table 5.4. Then, based on the properties of the formed graphs, graph theory principles can be applied in order to make selection based on the number of edges, and consequently, show that this specific subset of inputs or metrics is the commonly repeated, and therefore, the most important to be considered in the sustainable DC simulation.

The formed graph is a rooted tree, which is defined as a tree with a designated vertex called the root which represents the input, and each edge is implicitly directed away from the root, and going directly into the siblings, which represent the computed inputs or metrics, the edges of a rooted tree are assigned a natural orientation away from the root, in which case the structure becomes a directed rooted tree or out-tree. The formed tree in Figure 5.15 is a 12-ary tree, which consists of one vertex as a root and 10 children representing the computed inputs or metrics [GY05].

Figure 5.15 represents the rooted tree graph of the input "Total Facility Energy", it is a 12-ary tree with a root vertix, and it shows the relationship between this input which is represented as the parent vertix or the root, and the siblings which are the other inputs and metrics, for which it is used as input.

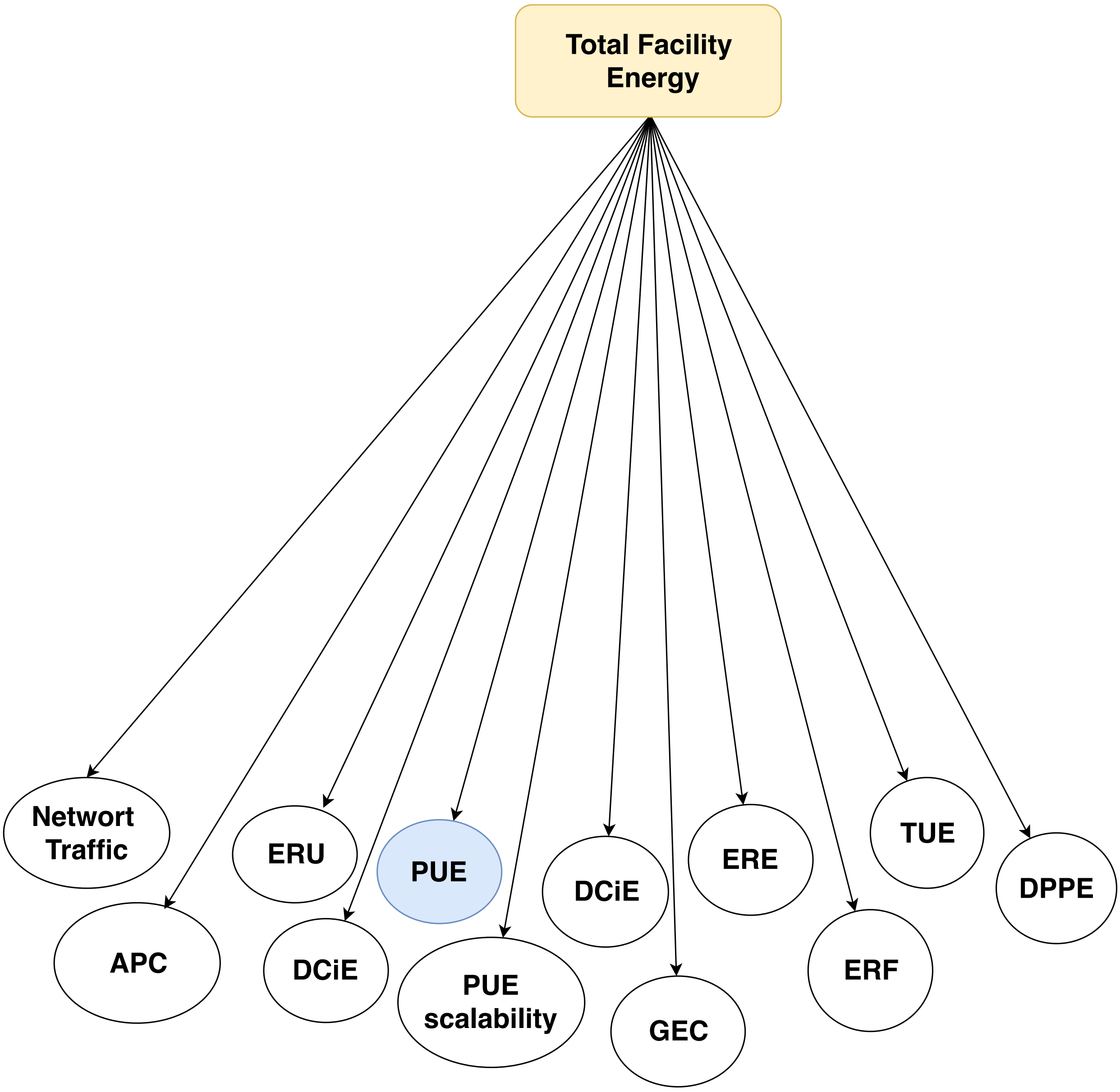


Figure 5.15: The rooted tree graph of the input "Total Facility Energy".

Figure 5.16 represents the rooted tree graph of the input "IT Energy", it is a 14-ary tree with a root vertex, and it shows the relationship between this input and the siblings which are the other inputs and metrics, for which it is used as input.

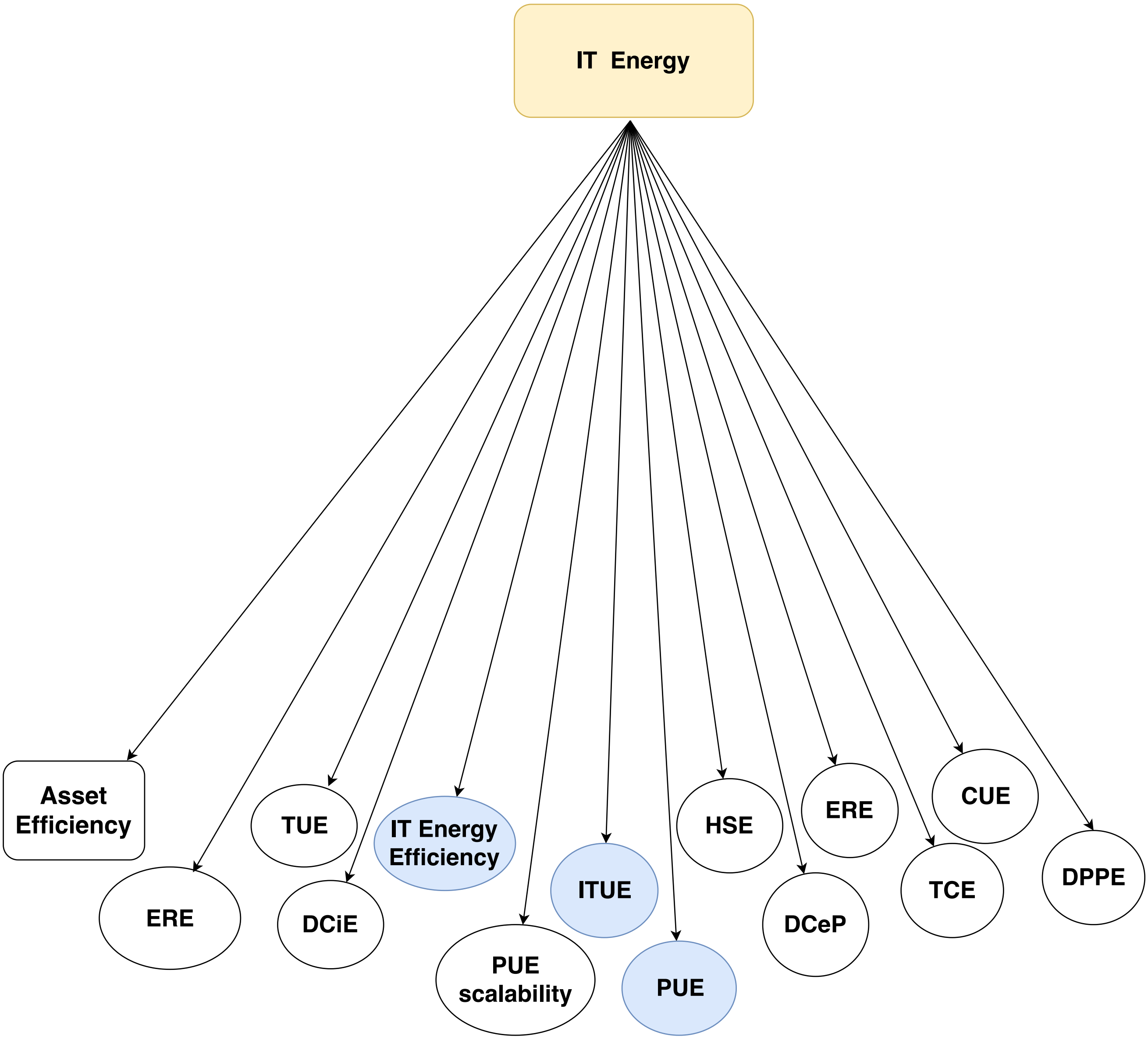


Figure 5.16: The rooted tree graph of "IT Energy".

Figure 5.17 represents the rooted tree graph of the input "Total Facility Power", it is a 7-ary tree with a root vertix, and it shows the relationship between this input and the siblings which are the other inputs and metrics, for which it is used as input.

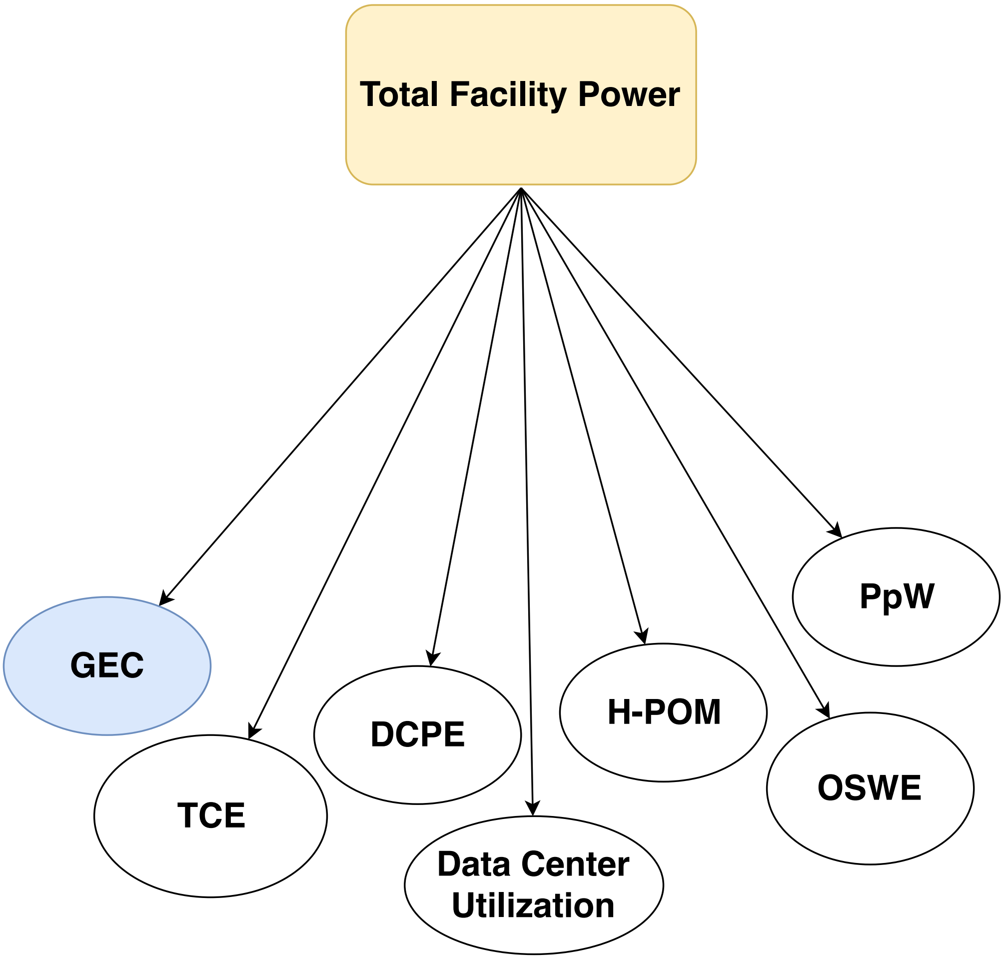


Figure 5.17: The rooted tree graph of "Total Facility Power".

Figure 5.18 represents the rooted tree graph of the inputs; “Peak power” and “Actual Power”, the peak power rooted tree is a 4-ary tree a root vertix represents the input, and the second tree shows the relationship between the actual power as an input and the siblings which are the other inputs and metrics, for which it is used as input.

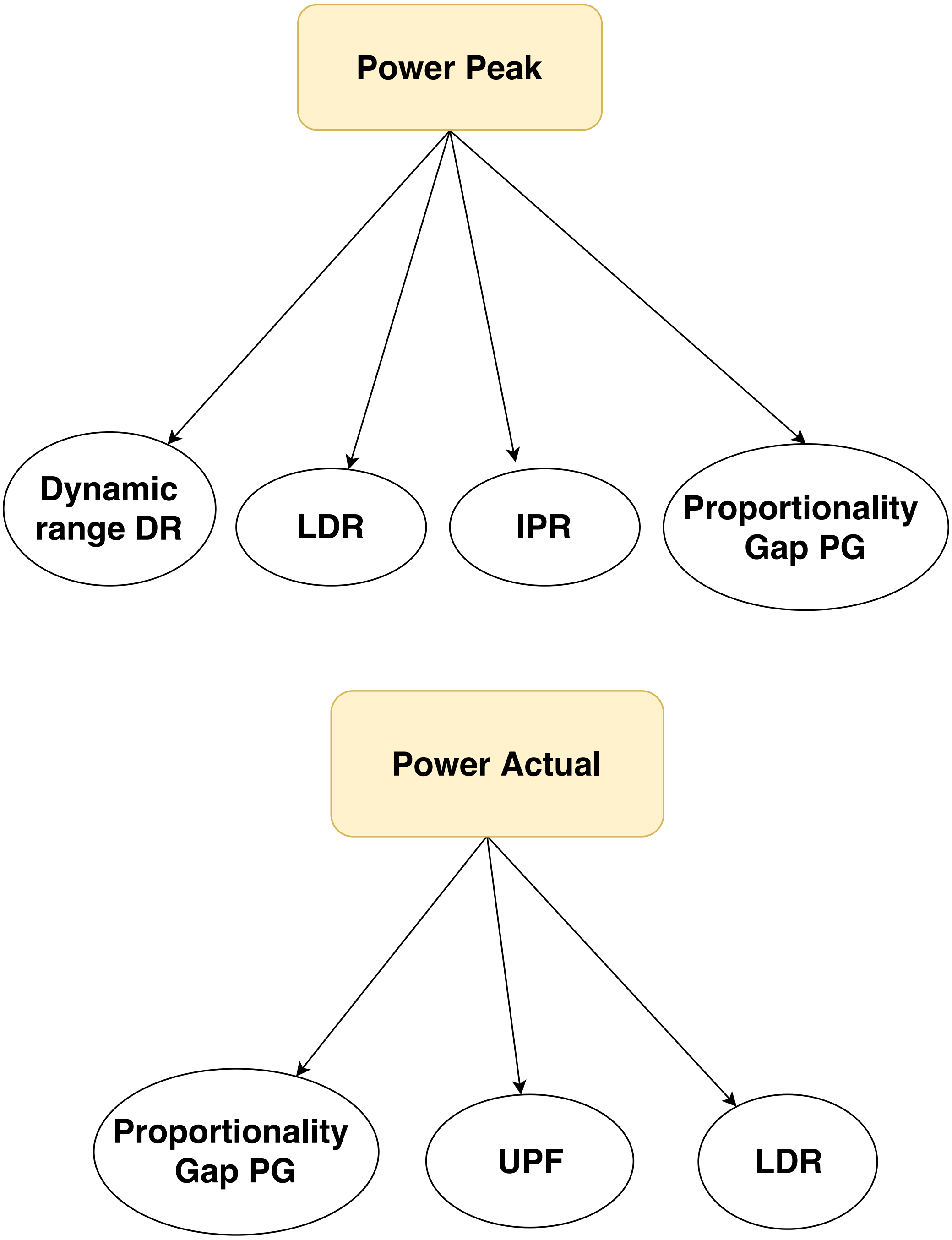


Figure 5.18: The rooted tree graph of “Actual Power” and “Peak Power”.

Figure 5.19 represents the rooted tree graph of the inputs; “EDC Baseline i” and “Useful Work”, the EDC Baseline rooted tree is a 4-ary tree a root vertix represents the input, and the useful work input is represented as a 6-ary tree, it shows the relationship between the useful work as an input and the siblings which are the other inputs and metrics, for which it is used as input.

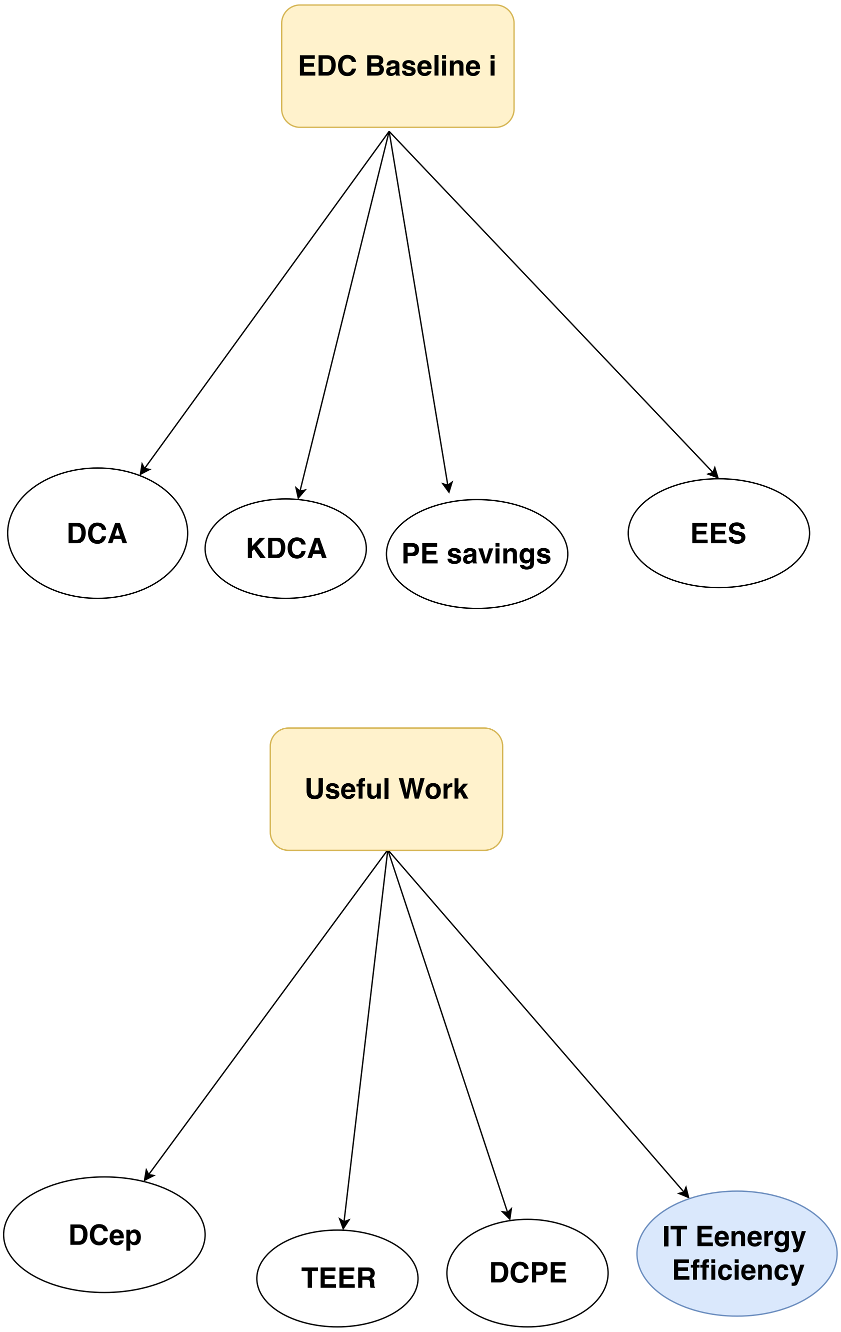


Figure 5.19: The rooted tree graph of “EDC Baseline i” and “Useful Work”.

Figure 5.20 represents the rooted tree graph of the metrics; “IT Equipment Utilization”, “Total Raw Memory Capacity” and “Reused Energy”. All of the three rooted tree are a 2-ary tree.

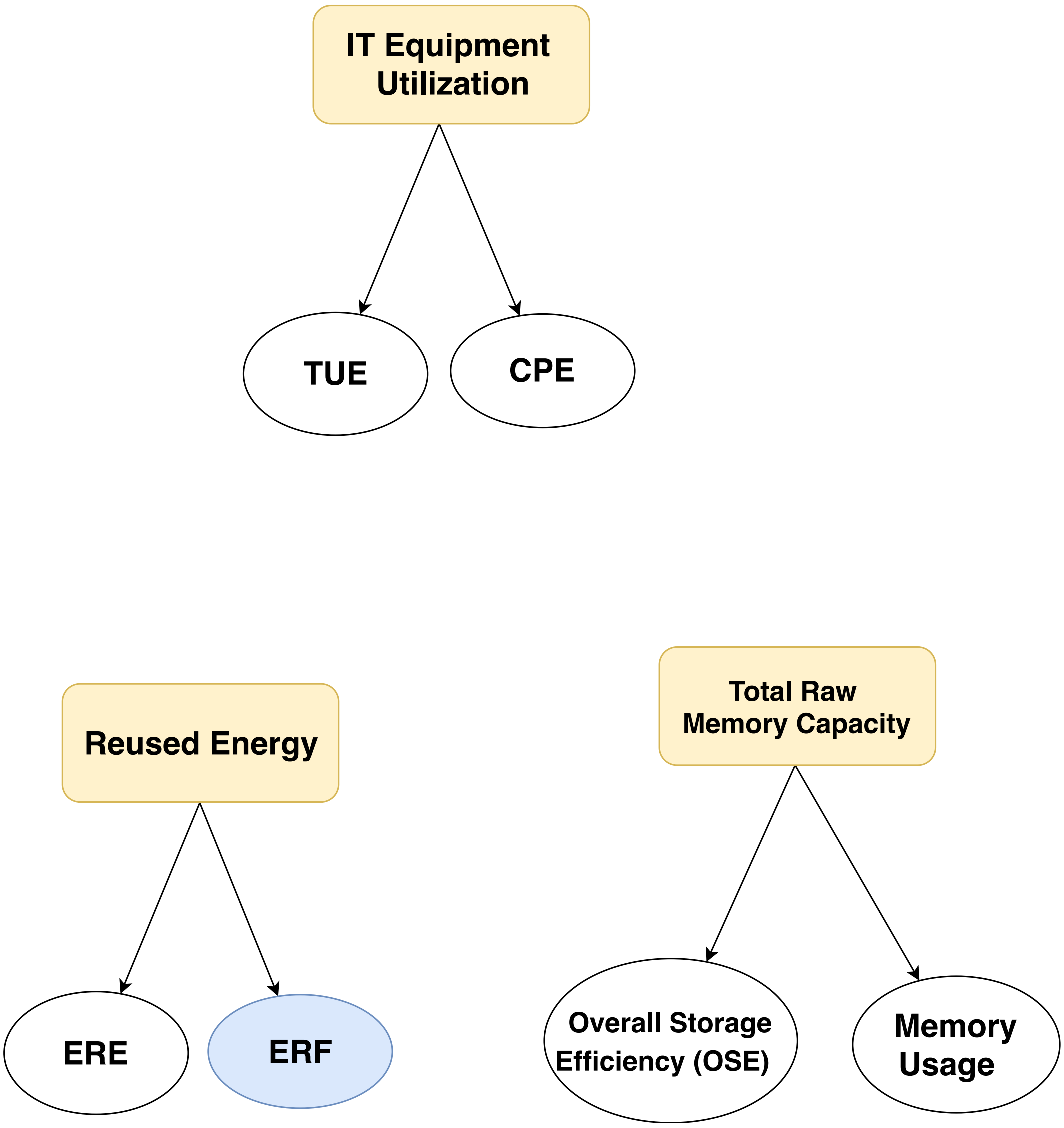


Figure 5.20: The rooted tree graph of “IT Equipment Utilization”, “Total Raw Memory Capacity”, and “Reused Energy”

Figure 5.21 represents the rooted tree graph of the metrics; “UPS”, “IT Energy Efficiency” and “CRAC Flow”. The UPS rooted tree is a 4-ary tree a root vertix represents the commonly reused metric, and the ITEE tree shows the relationship between IT energy efficiency as a metric and the siblings which are the other inputs and metrics, for which it is used as input. And the CRAC Flow rooted tree that shows that this input is used to calculate 2 metrics.

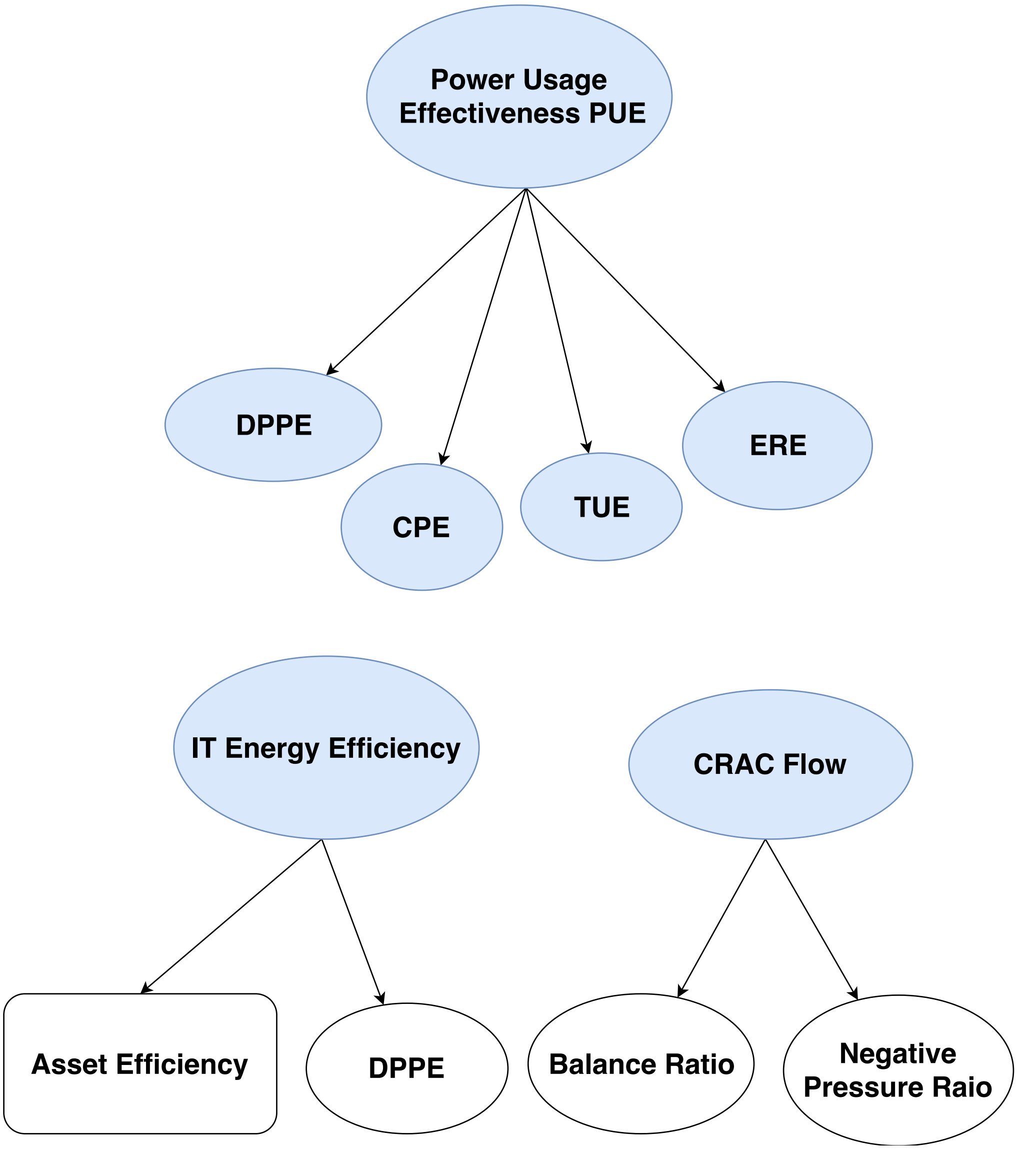


Figure 5.21: The rooted tree graph of “UPS” and “IT Energy Efficiency”.

To sum up, the rooted tree graphs [GY05] were created to show that the commonly reused inputs and metrics are dominant vertices, and have higher importance in their graphs.

5.2 Results

The rooted tree graphs were created to show that, the commonly reused inputs and metrics have higher importance since they are commonly reused many times in order to evaluate the sustainability metrics. Therefore, the change in these inputs lead to significant change in the evaluated metrics. The sustainable DC simulator will be used to monitor this change. Table 5.3 lists the final list of commonly reused inputs and metrics based on their frequency of contribution in evaluating the sustainability metrics as illustrated by the rooted tree graphs.

Table 5.3: The final list of commonly reused inputs and metrics.

List of Inputs	List of Metrics
IT Energy	Power Usage Effectiveness (PUE)
Total Facility Energy	IT Energy Efficiency (ITEE)
Total Facility Power	CRAC Flow (Mc)
IT Power	IT Usage Effectiveness (ITUE)
Useful Work	Server Compute Efficiency (SCE)
EDC Baseline i	Green Energy Coefficient (GEC)
Actual Power	Energy Reuse Factor (ERF)
Peak Power	
Used Memory by Servers or Application	
Planned Energy EPi	
IT Equipment Utilization ITEU	
Total Deployed Servers	
Total DC space	
Performance	
Cost ei current	
E oth DCi base	
EDC i current	
EDC Real i	
Idle Power	
Reused Energy	
Ideal Power	
Server Air Flow	
Total Raw Memory Capacity	
Total Energy Delivered to Components	

Table 5.3 is considered the input base, from which the essential sustainability metrics are expected to be computed in the sustainable DC simulator. It is necessary to check the list of inputs and metrics in Table 5.3 to determine which metrics can be calculated out of these inputs.

The sustainable DC simulator requires the minimum number of commonly reused inputs in order to compute the maximum number of available metrics. We determine the set of metrics that can be calculated using these commonly reused inputs by checking the inputs in table 5.3 and find out which metrics can be evaluated using these set of inputs. The commonly reused inputs in previous table are analysed, and Table 5.4 shows a summary of the maximum number of sustainability metrics that can be computed using this minimum number of commonly reused inputs.

Table 5.4: Sustainability metrics and required inputs

Metrics	Required Inputs
Power Usage Effectiveness (PUE)	Total Facility Energy IT Energy
Data Center Infrastructure Efficiency (DCiE)	Total Facility Energy IT Energy
Carbon Usage Effectiveness (CUE)	PUE Carbon Emission Factor (given according to the DC location)
Storage Usage	Total Memory Capacity Used Memory
Data Center Performance Efficiency (DCPE)	Useful Work Total Facility Power
Data Center Energy Productivity (DCeP)	Useful Work Total Facility Energy
Energy Reuse Effectiveness (ERE)	Total Energy Reused Energy IT Energy
Energy Reuse Factor (ERF)	Total Facility Energy Reused Energy
Data Center Productivity (DCP)	Useful Work Total Facility Power
Idle-to-peak Power Ratio (IPR)	Idle Power Peak Power
Dynamic Range (DR)	Idle Power Peak Power
IT-Power Usage Effectiveness (ITUE)	IT Energy Total Energy Delivered to Compute Components
Balance Ratio	Server Air Flow CRAC Flow
Total-Power Usage Effectiveness (TUE)	PUE IT-Power Usage Effectiveness (ITUE)

Continued on next page

Table 5.4 – Continued from previous page

Metrics	Required Inputs
Compute Power Efficiency (CPE)	IT Power IT Equipment Utilization Total Facility Power
Data Center Utilization (UDC)	IT Power IT Equipment Utilization

6 Data Center Simulation

In this chapter, we provide a brief description about each of the most relevant simulation platforms that can be potentially used for simulating a DC according to our analysis results from the previous chapter. The simulators are compared to know if there is one simulator that can simulate the minimum number of commonly reused inputs and evaluate the maximum number of metrics.

6.1 Existing Simulation Platforms

Following are some of the major platforms or tools that could be potentially used to simulate a sustainable DC. Some of the simulators are web-based applications, which provide an interactive interface to simulate a DC, and the others are source codes that have to be compiled and run to see the results.

A Multi-tier Data Center Simulation Platform (MDCSim)

MDCSim is designed to capture all the important design specifics of the underlying communication paradigm, kernel level scheduling artifacts, and the application level interactions among the tiers of a three-tier DC since it is designed as a pluggable three-level architecture. MDCSim flexibility is attributed to its ability in experimenting with different design alternatives in the three layers, and in analyzing both the performance and power consumption with realistic workloads. The scalability of the simulator is demonstrated with analyses of different DC configurations. This simulator is quite accurate in estimating the throughput, response time, and power consumption parameters. The paper demonstrated configuration analysis for performance optimization in terms of reduced network latency using the simulation platform. Moreover, a methodology to perform power measurement in a multi-tier DC using MDCSim was provided [LSN+09].

Data Center Simulator (DCSim)

DCSim (Data Center Simulator) an an easily customized and extensible simulation tool for simulating a multi-tenant, virtualized DC. Furthermore, it provides an application model that can simulate the interactions and dependencies between many VMs working together to provide a single service, such as in the case of a multi-tiered web application. It supports other features of virtualization, such as a work conserving CPU scheduler used in modern hypervisors, resource allocation, and VM migration and replication must also be available. Moreover, it allows host power states (on, off, suspended) to be modelled with appropriate transition times between states. DCSim was designed specifically to study VM management in a DC providing an IaaS Cloud [TKBL12].

DCSim supports Virtualisation modeling which is not supported by many of the other simulators. One advantage of DCSim over GreenCloud, MDCSim, and GDCSim is that it focuses on a virtualized DC, which helps in providing IaaS to multiple tenants, similar to CloudSim. One other advantage of DCSim over CloudSim is that it focuses on transactional, continuous workloads. Furthermore, DCSim provides the additional capability of modelling replicated VMs sharing incoming workload as well as dependencies between VMs that are part of a multi-tiered application. Similarly, Service Level Agreement (SLA) achievement can also be more directly and easily measured and available to management elements within the simulation. It generates a log containing details of the simulation as it progressed, including metric values at regular time intervals, host and VM states, and other events. It can be used to generate statistics, graphs, or other visualizations in order to fully understand the behaviour of the simulated DC and management algorithms. Figure 6.1 shows the general architecture of DCSim, which simulates a virtualized DC operating an Infrastructure as a Service cloud. A DC consists of a set of interconnected Hosts, governed by a set of Management Policies. Each host has a set of Resource Managers that manage local resource allocation and a CPU scheduler which decides the execution procedure of VMs, and a power model which models the power consumption by the host at any time.

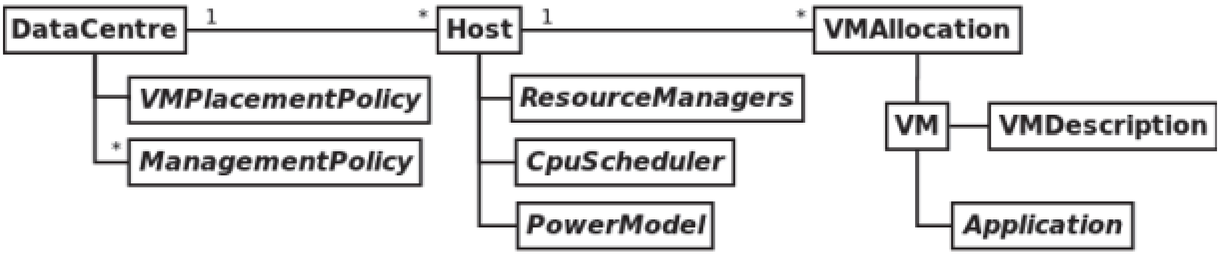


Figure 6.1: The general architecture of DCSim [TKBL12]

Following are a set of metrics, which are computed by DCSim to determine the behaviour of the DC during the simulation process [TKBL12]:

SLA Violation: When a VM requires more resources than is available to it, the VM will experience performance degradation. DCSim considers this to be an SLA violation, and the percentage of CPU resource required by running VMs but not available is recorded.

DC Utilization: The overall utilization of the DC is calculated as the percentage of total CPU capacity in the DC that is currently in use.

Active Hosts: DCSim records the minimum, maximum, and average number of hosts that are On at any given time during the simulation.

Host-hours: Host-hours is the combined total of the active time of every host in the simulation. That is, if 10 hosts were active for 30 simulation minutes each, then 5 host-hours were used. This gives a combined measure of the number of hosts that were required to meet the workload demand throughout the entire simulation run.

Active Host Utilization: DCSim measures the CPU utilization of all hosts that are currently in the On state. The higher the average utilization, the more efficiently resources are being used.

Number of Migrations: The number of VM migrations triggered during the simulation is recorded.

Power Consumption: Power consumption is calculated for each host, and the total kilowatt-hours consumed during the simulation are reported.

CloudSim

CloudSim is developed to make it easier to measure the performance of resource allocation policies and scheduling algorithms in details in cloud environments for various applications and service models under different constraints of varying load, power consumption, heat dissipation and system size is a daunting task to tackle. It is new generalized and extensible simulation framework that enables seamless modelling, simulation, and experimentation of emerging Cloud computing infrastructures and management services [CRDB09]. It supports the modelling and instantiation of large scale Cloud computing infrastructure, including DCs on a single physical computing node and java virtual machine.

Furthermore, it is considered a self-contained platform for modelling DCs, service brokers, scheduling, and allocations policies. It provides the availability of virtualization engine, which aids in the creation and management of multiple, independent, and co-hosted virtualized services on a DC node. It supports also the flexibility to switch between space-shared and time-shared allocation of processing cores to virtualized services. A DC can be modeled using CloudSim as the core infrastructure level services (hardware, software) offered by resource providers in a Cloud computing environment. It encapsulates a set of compute hosts (blade servers) that can be either homogeneous or heterogeneous as regards to their resource configurations (memory, cores, capacity, and storage). Every DC component instantiates a generalized resource provisioning component that implements a set of policies for allocating bandwidth, memory, and storage devices [CRDB09]. Figure 6.2 shows the flow of communication among the main CloudSim entities. Firstly, each DC entity registers itself with the cloud information service registry. Then, this service provides a database for mapping user requests to relevant cloud providers.

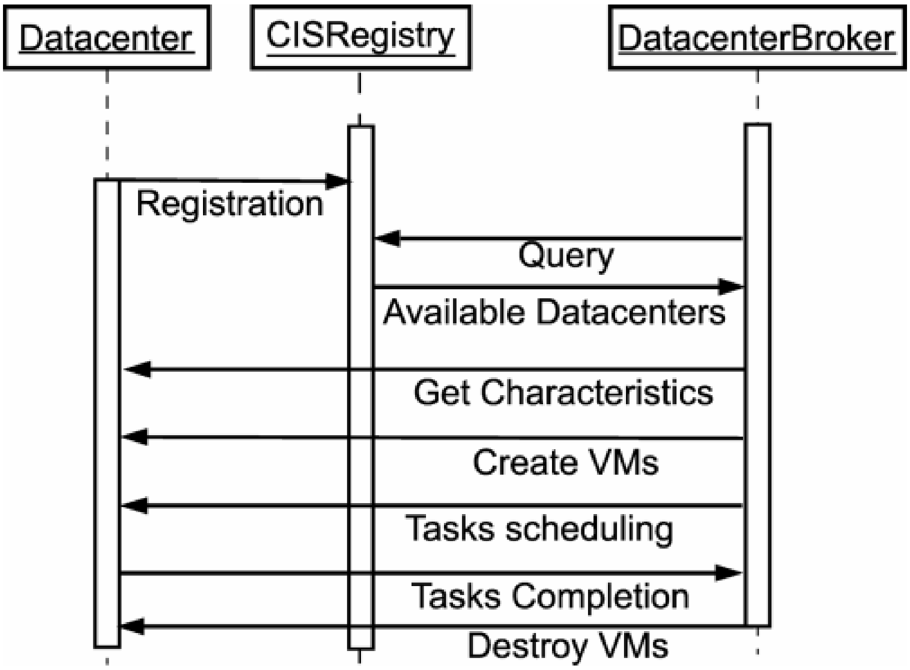


Figure 6.2: Data flow in CloudSim [CRDB09]

CloudNetSim++

CloudNetSim++ is a modeling and simulation toolkit to facilitate simulation of distributed DC architectures, energy models, and high speed DC communication network. It is designed to allow researchers to incorporate their custom protocols and applications to analyze under realistic DC architectures with network traffic patterns, as well as analyzing energy consumption by varying number of nodes and other parameters, in addition to standard network performance measures like delay and throughput for various topologies.

It offers support for a heterogeneous computing environment in a way that makes the user can define resources with varying CPU power, Random Access Memory (RAM), and storage space for individual compute nodes. Besides providing support for multiple users, delivering a more realistic multi-tenant cloud environment, it computes energy consumed by network elements, such as switches and routers. It also allows the user to specify VMs with high priority, to deal with and service high priority traffic, as well as providing different pricing modules for DCs services.

It is the first cloud computing simulator that uses real network physical characteristics to model distributed DCs. It provides a generic framework that allows users to define a service level agreement policy, scheduling algorithms, and modules for different components of DCs without worrying about low level details with ease and minimum effort. The paper also shows the flexibility and effectiveness of CloudNetSim++ through experimental results demonstrated using real-world DC workloads. More importantly, it provides a rich Graphical User Interface (GUI) based environment that allows users to visualize the entire network along with fine-grained packet flows, this GUI supports further simplified analysis and debugging [MBA+14].

GroudSim

GroudSim is a grid and cloud simulation tool intended for scientific applications based on a scalable simulation-independent discrete-event core. It provides a comprehensive set of features for complex simulation scenarios from simple job executions on leased computing resources to calculation of costs, and background load on resources. These simulations can be parameterised and are easily extendable by probability distribution packages for failures which normally occur in complex environments. Experimental results demonstrate the improved scalability of GroudSim compared to a related process-based approach. It supports finding valid solutions for the issues of resource management, scheduling, fault tolerance, or Quality of Service, which require a lot of of simulation experiments [OPPF10].

Parasol

Parasol is a prototype green DC that was built mainly as a research platform. It comprises a small container, a set of solar panels, a battery bank, a free cooling unit, a direct-expansion air conditioner and a grid-tie. Moreover, [GKL+13] provides the green switch, which is a model-based framework for managing workloads and energy sources in green DCs. Green DCs which partially or completely powered by renewable energy that will either generate their own renewable energy or draw it directly from an existing nearby plant. In addition to reducing carbon footprints, renewable energy can potentially reduce energy costs, reduce peak power costs, or both.

Nevertheless, certain renewable fuels are intermittent, which requires approaches for tackling the energy supply variability. One approach is to use batteries or the electrical grid as a backup for the renewable energy. It may also be possible to adapt the workload to match the renewable energy supply. For highest benefits, green DC operators must intelligently manage their workloads and the sources of energy at their disposal, and this is the reason that led to provide Parasol and green switch which together can help in managing the workload [GKL+13].

Green Data Center Simulator (GDCSim)

GDCSim is an integrated tool chain for analyzing green DC physical design and resource management techniques, a study was provided in [GGB+11], it testes different DC geometries, workload characteristics, platform power management schemes, scheduling algorithms and DC configurations to build this holistic simulation tool, which supports automated processing that allows interfacing different modules together and does not require user intervention once the simulation has started, besides the online analysis capability, which allows real time simulation of management decisions based on changes in the physical environment in the DC. Figure 6.3 shows the architecture of the GDCSim tool.

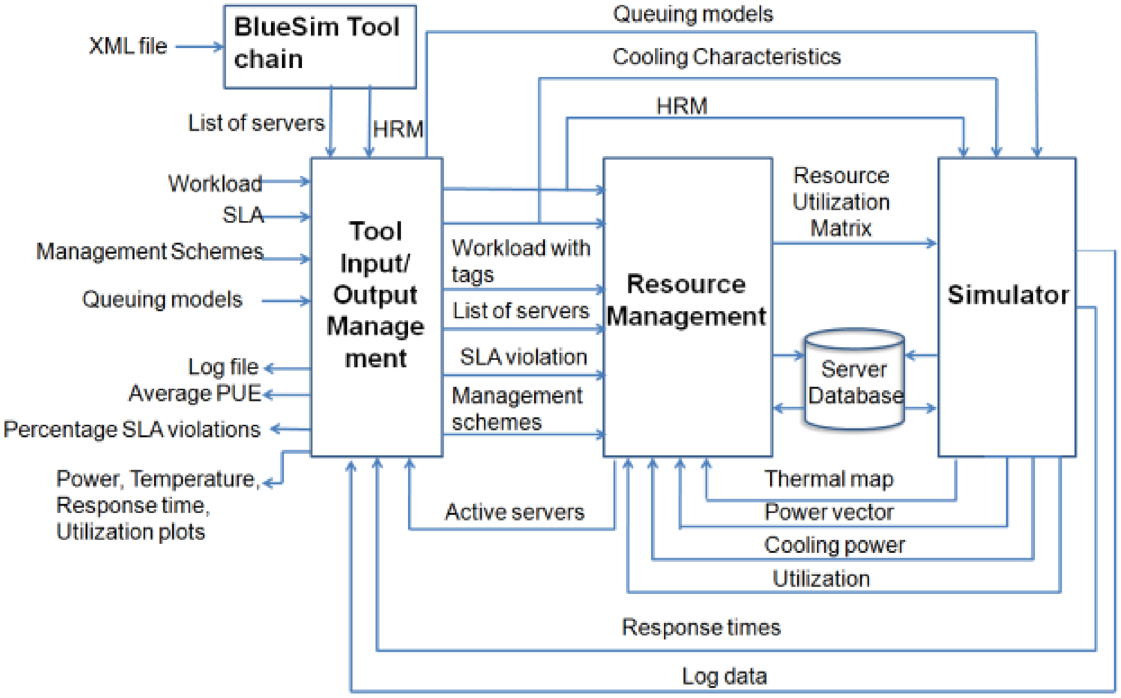


Figure 6.3: Architecture of GDCSim [GGB+11]

It supports iterative design analysis, which enables design time testing and analysis of different DC configurations before deployment, it supports also thermal analysis capability, which characterizes the thermal effects within the DC room at a given time, but it doesn't give any thermal feedbacks to the management of the DC. One of the most important features of this tool is workload and power management, which enables workload scheduling and controlling the power modes of servers for large DCs. It enables feedback of information on temperature and air flow patterns in the DC to the management algorithms and the closed loop operation of the servers and cooling units (CRAC) to achieve energy efficient operation, which is a crucial point towards achieving greenness in DCs. This tool can be used in conjunction with any resource management algorithm to estimate DC performance. Figure 6.3 shows that, the characteristics of a DC layout are provided through an XML file, which provides an easy and intuitive way of DC physical modeling. Additionally, the tool facilitates testing of online resource management algorithms on different DC designs and workload types. The result from the case studies, which have been applied by the authors, shows that thermal aware workload management schemes avoid redlining of servers while improving the efficiency. Further, the tool captures the transient behavior of the DC which results in improved accuracy over the analysis [GGB+11].

Autonomic Data Center Management Simulator (ADCMSim)

ADCMSim is a platform to evaluate different collaboration models between different autonomic computing agents and to explore the influence of different collaboration architectures on energy consumption of a DC while considering compliance with SLAs, the DC in this simulator is represented as an entity managed by autonomic agents which make use of defined policies, that are defined atvarious levels, for the purpose of managing a lot of aspects of a DC, from allocation of jobs to management or workload [NB12].

ADCMSim supports the ability to support different types of servers with different levels of energy consumption and computing power, moreover, it allows defining several types of workloads, such as, compute intensive or interactive jobs. Furthermore, it supports dynamic resource allocation for dynamically allocating server to meet different application requirements under changing workloads, as well as supporting different workload scheduling algorithms. Moreover, it has the ability to define different kinds of SLA for each type of workload and also keep tracking of SLA violations and the ability to record energy consumption of the DC and it can manage the CPU working frequency of servers [NB12].

SimScale

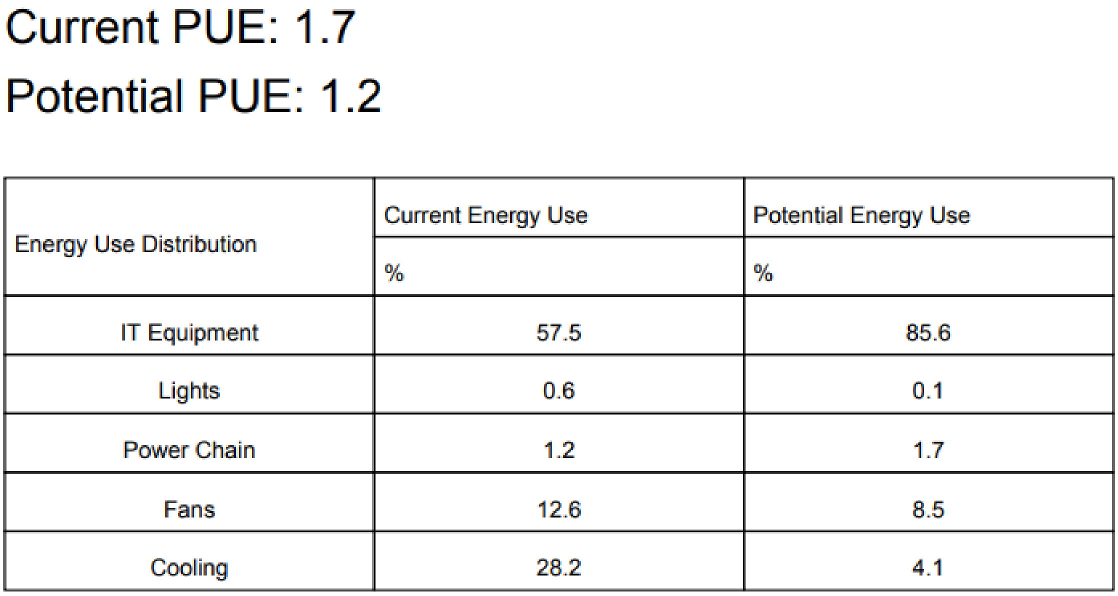
SimScale is an online cloud-based platform for the Computational Fluid Dynamics (CFD) analysis, it offers a higher level of computing power as compared to a local system, computational fluid dynamics simulates fluid motion and heat transfer using numerical approaches. Simscale CFD software can analyze a range of problems related to laminar and turbulent flows. More over, it can reduce energy consumption for HVAC systems via better heat exchanger designs, ducting, fan or cooling unit placement and get energy efficiency certification for DCs, which is useful for designing a sustainable DC [Simb].

SimScale helps engineers to develop products that comply with current energy and power quality standards and regulatory provisions. It can improve the thermal management of a DC, heat sinks or engines. Furthermore, it can optimize electronics cooling on several levels from chip to room, predict resulting temperatures in product's case, reduce failure risk and decide on fan placement, sizing, and speed for cost-efficient cooling [Simb].

The Center of Expertise for Energy Efficiency in Data Centers

Since DCs are energy-intensive facilities and potential benefits of DC energy efficiency include 20-40% typical savings, the center of expertise for energy efficiency helps federal agencies and other organizations implement DC energy efficiency projects by providing a set of powerful tools, best practices and analyses. The center online platform provides the DC Profiler (DC Pro) Tools which are used for early stage power assessment, this tool helps DC operators estimate PUE, the industry standard for understanding and improving the energy efficiency of DC infrastructure systems. Current DC Pro Tools include DC Pro, which estimates PUE and provides tailored recommendations for improvement. Moreover, it provides a set of excel-based worksheet to document metrics, actions, and measurements from DC assessments. Additionally, it has two excel-based tools designed to optimize air management in DCs while enhancing energy efficiency. Both tools were intended to help users accelerate the energy savings in DCs without affecting the thermal IT equipment environment [Eneb]. Additionally, it provides a set of excel sheets which help in documenting the metrics, actions, and measurements from DC assessments. The toolkit provides the DC Pro tool which is used for the first step of creating a DC simulator by entering the initial inputs of the DC.

For clarification, a sample DC simulator is created using DC Pro. Once all the inputs are given, the tool provides a table which lists the current energy usage and the potential energy usage that can be achieved. Moreover, it calculates the current PUE and anticipates applying improvements to achieve the given potential PUE as shown on figure 6.4.



Current PUE: 1.7

Potential PUE: 1.2

Energy Use Distribution	Current Energy Use	Potential Energy Use
	%	%
IT Equipment	57.5	85.6
Lights	0.6	0.1
Power Chain	1.2	1.7
Fans	12.6	8.5
Cooling	28.2	4.1

Figure 6.4: Energy results of a sample DC simulation using DC Pro [Enea]

Figure 6.5 illustrates how DC Pro visualizes the energy usage distribution. It shows that 57% of the power is dedicated for the IT equipment and 28% is for cooling. Based on the visualized result of energy distribution in the DC, DC Pro provides a set of tailored recommendations on how to increase energy efficiency.

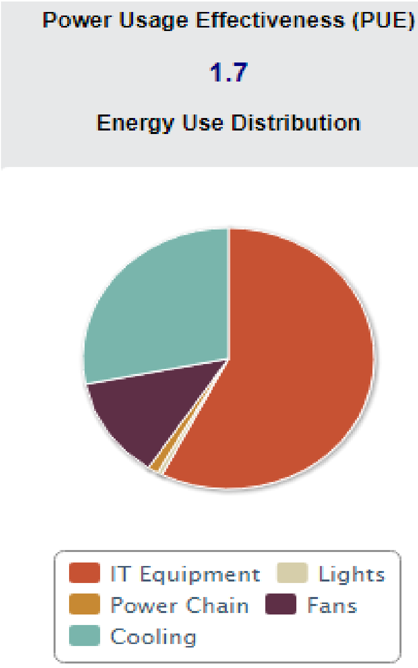


Figure 6.5: Power distribution visualisation of a sample DC simulation using DC Pro [Enea]

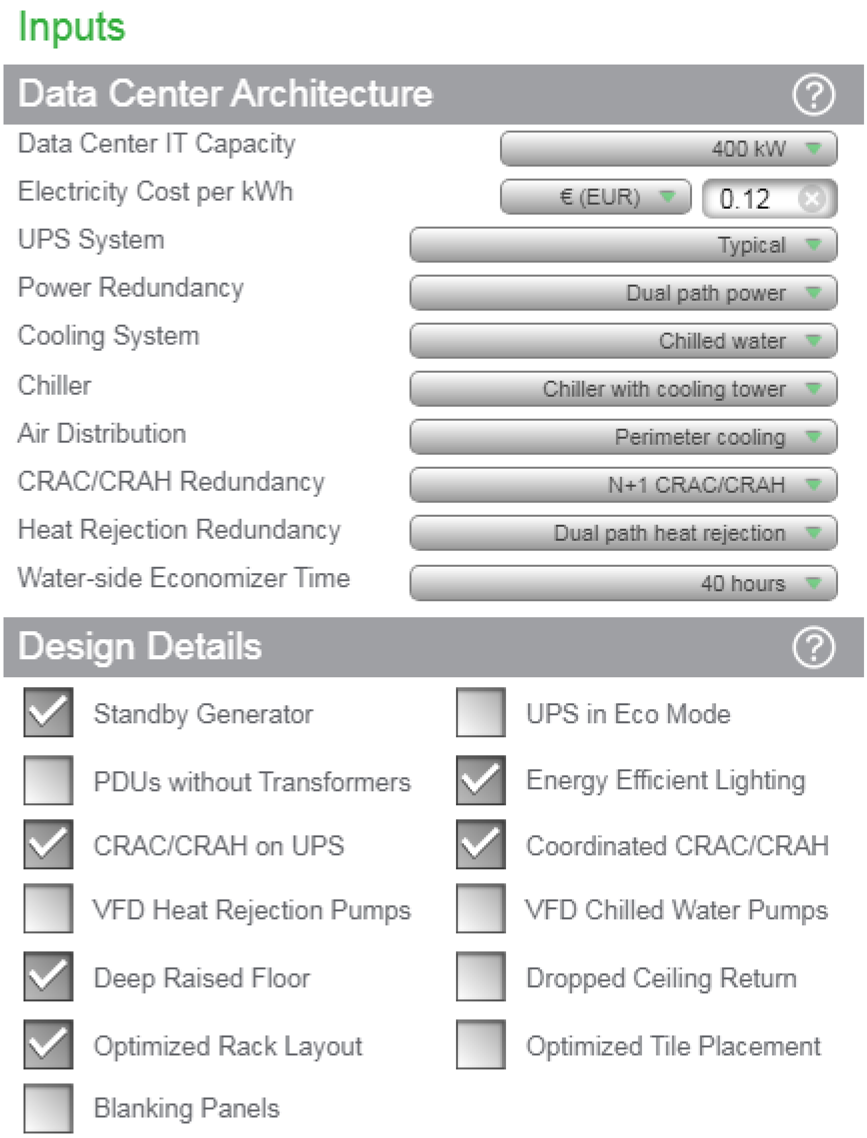
Schneider Electric

Schneider Electric provides energy and automation digital solutions for efficiency and sustainability in DCs. It focuses on risk mitigation in all operational and maintenance activities and work procedures. Schneider Electric contributes to DC efficiency as it provides a set of web-based tools that allow giving inputs and evaluating some of the relevant DC sustainability metrics [Eled].

Data Center Efficiency (PUE Calculator) is one of these tools, which is intended to help DC operators estimate what they can expect for their annualized PUE, and helps them quantify how much PUE can improve when you take specific actions or select different architecture choices. the input parameters include DC size, expected IT load, power cooling architecture, and other selectable DC characteristics. The simulation results show the overall PUE, and several charts, including PUE Data Center Productivity (DCP) curve at varying loads, a breakdown of the power cooling losses, and breakdowns of the energy costs [Eleb].

This tool shows how design decisions affects the efficiency a DC. As the DC operator inputs details regarding the power and cooling configuration, results are calculated. Actual DCs can vary greatly from this model due to the efficiency and sizing of the actual power and cooling devices used. The primary value of this tool is to evaluate the comparative effects of different design or condition changes to obtain actual efficiency values for a specific DC [Elec].

Figure 6.6 shows the set of inputs that can be provided, such as electricity cost, cooling system type, power redundancy, air distribution, and DC IT capacity which is the maximum kW of IT load that can be supported by a DC. Moreover, some design details can be provided in order to obtain accurate results.



Inputs

Data Center Architecture

Data Center IT Capacity	400 kW
Electricity Cost per kWh	€ (EUR) 0.12
UPS System	Typical
Power Redundancy	Dual path power
Cooling System	Chilled water
Chiller	Chiller with cooling tower
Air Distribution	Perimeter cooling
CRAC/CRAH Redundancy	N+1 CRAC/CRAH
Heat Rejection Redundancy	Dual path heat rejection
Water-side Economizer Time	40 hours

Design Details

☑ Standby Generator

☐ PDUs without Transformers

☑ CRAC/CRAH on UPS

☐ VFD Heat Rejection Pumps

☑ Deep Raised Floor

☑ Optimized Rack Layout

☐ Blanking Panels

☐ UPS in Eco Mode

☑ Energy Efficient Lighting

☑ Coordinated CRAC/CRAH

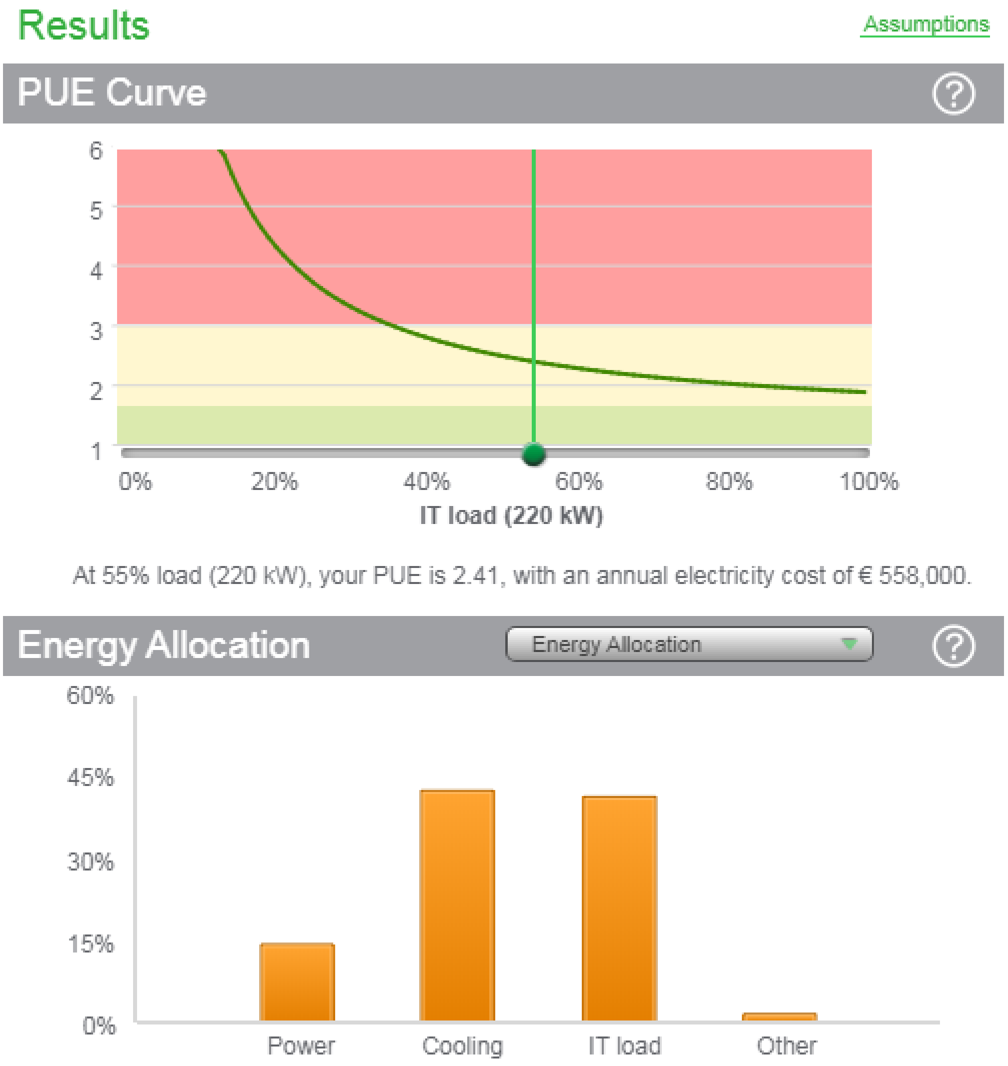
☐ VFD Chilled Water Pumps

☐ Dropped Ceiling Return

☐ Optimized Tile Placement

Figure 6.6: Data center inputs and design details in PUE calculator [Elec]

Figure 6.7 shows the PUE curve and the energy allocation chart, which shows the breakdown of energy consumption. It shows how cooling consumes a large amount of energy.



At 55% load (220 kW), your PUE is 2.41, with an annual electricity cost of € 558,000.

Figure 6.7: Results of PUE calculator [Elec]

6.2 Sustainable Data Center Simulation Requirements

We need a DC simulator which can make use of the set of commonly reused inputs and evaluate as much metrics as possible. In order to decide which simulator can be used, it was required then to compile and run the source codes of the open source platforms to check how they work, and if it is possible to use or extend one of them. But the results show that most of the open source tools focus essentially on virtualisation and automation issues and they lack the possibility of simulating the commonly reused inputs from Table 5.4. On the other hand, web-based simulators are rich with capabilities such as interactive simulation modeling and visual assessments.

Once the existing DC simulators are listed and analysed, we can then define the requirements of the intended sustainable DC simulator. The main rule of this simulator is to evaluate the maximum number of sustainability metrics from the minimum number of commonly reused inputs. Table 5.4 lists the 16 sustainability metrics that can be evaluated using 15 different commonly reused inputs. The potentially evaluated metrics should reflect the energy efficiency, greenness, productivity, performance, airflow, and storage usage of a DC. These requirements are defined in Table 6.1, which shows how the other simulators fulfill these requirements.

Table 6.1: Comparison between the simulators and the requirements

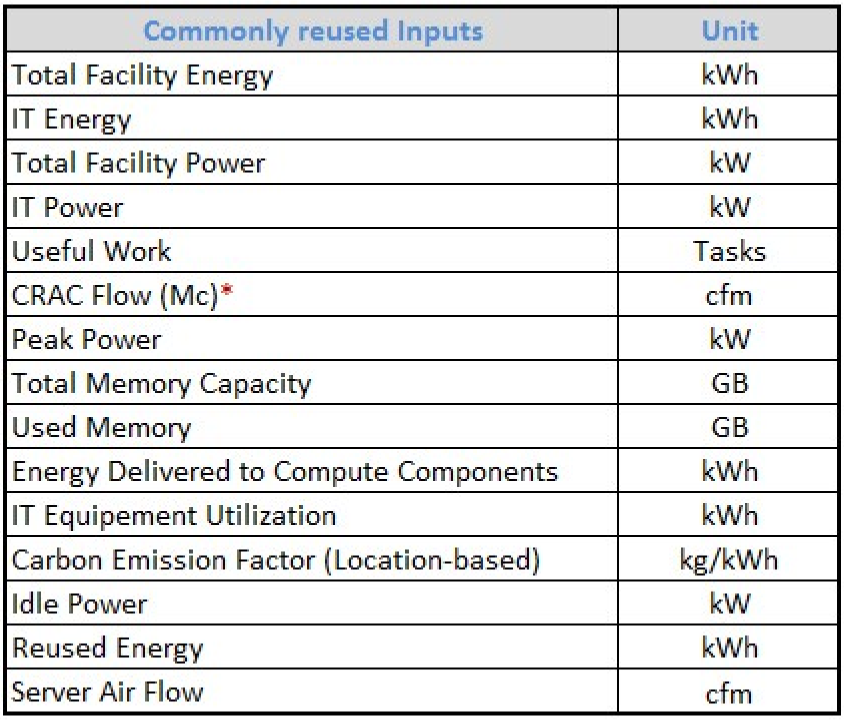
Requirements	**DCSim**	**CloudSim**	**GDCSim**	**CoE**	**Schneider Electric**
Energy Efficiency Metrics	Yes	Yes	Yes	Yes	Yes
Greenness Metrics	No	No	No	NO	Yes
Performance / Productivity Metrics	No	No	No	Yes	No
Air Management Metrics	No	No	Yes	Yes	Yes
Storage Metrics	No	No	No	Limited	No
GUI Support	No	Limited	No	Yes	Yes

Although Schneider Electric trade-off tools fulfill most of our requirements, but it the metrics evaluation is done separately on different calculators and there is no holistic application to support the DC simulation. Table 6.1 shows that DCSim and Cloudsim support only one of our requirements. On the other hand, GDCSim and Center of Experience toolkit support 2 and 4 requirements respectively. We conclude from the comparison that we have to build our own sustainable DC simulator, which can fulfill all of the requirements in one integrated tool.

6.3 Sustainable Data Center Simulator

Comparing the different DC simulators in Table 6.1 shows that, there is no single simulator which can support 5 important categories of sustainability metrics, which are energy efficiency, greenness, performance and productivity, air management, and storage metrics. Therefore, we create our own sustainable DC simulator which provide results related to the 5 aforementioned metrics categories. Moreover the simulator can make use of 15 commonly reused inputs to evaluate 16 sustainability metrics.

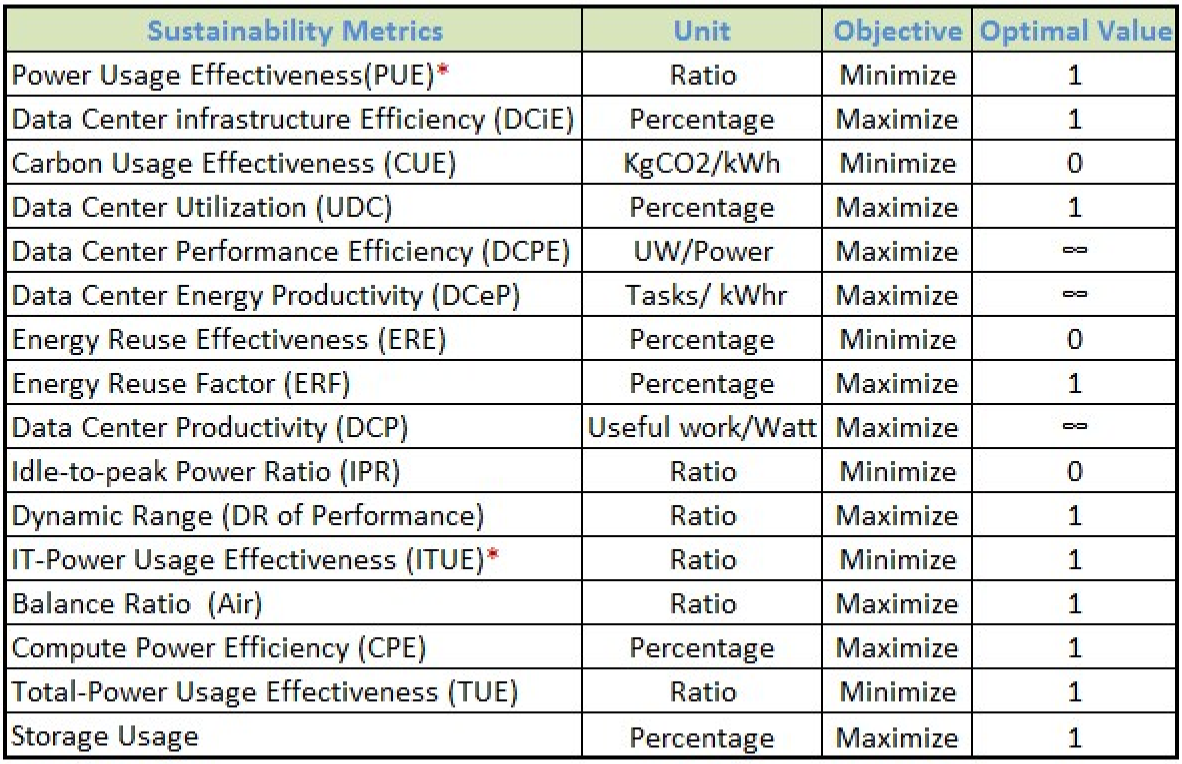
The sustainable DC simulator is an excel-based tool designed to help DC operators assess 16 different sustainability metrics. Especially, energy efficiency and greenness metrics can greatly help DC operators to identify the sustainability state of DCs, and accordingly they can form insights about improvement decisions that can be applied to to save energy and reduce carbon dioxide emissions, and consequently, have more efficient and sustainable DCs. Figure 6.8 shows the 15 commonly reused inputs that are used in the simulator.



Commonly reused Inputs	Unit
Total Facility Energy	kWh
IT Energy	kWh
Total Facility Power	kW
IT Power	kW
Useful Work	Tasks
CRAC Flow (Mc)*	cfm
Peak Power	kW
Total Memory Capacity	GB
Used Memory	GB
Energy Delivered to Compute Components	kWh
IT Equipement Utilization	kWh
Carbon Emission Factor (Location-based)	kg/kWh
Idle Power	kW
Reused Energy	kWh
Server Air Flow	cfm

Figure 6.8: The inputs of the sustainable DC simulator

Figure 6.9 shows the 16 sustainability metrics that are intended to be evaluated, the objective goal of each metric is given in the objective column and the optimal values are also provided. The metrics PUE and IT Power Usage Effectiveness (ITPUE) are the only reused metrics, because the metric ITPUE requires the two metrics to be evaluated. Moreover, the metric Carbon Usage Effectiveness (CUE) requires PUE to be evaluated. Therefore, these metrics are denoted with an asterisk to let the user know that these metrics have to be evaluated first, then the metrics ITPUE and CUE are evaluated accordingly. PUE and ITUE are then denoted with red asterisk, which means that these two metrics are required to calculate other metrics.

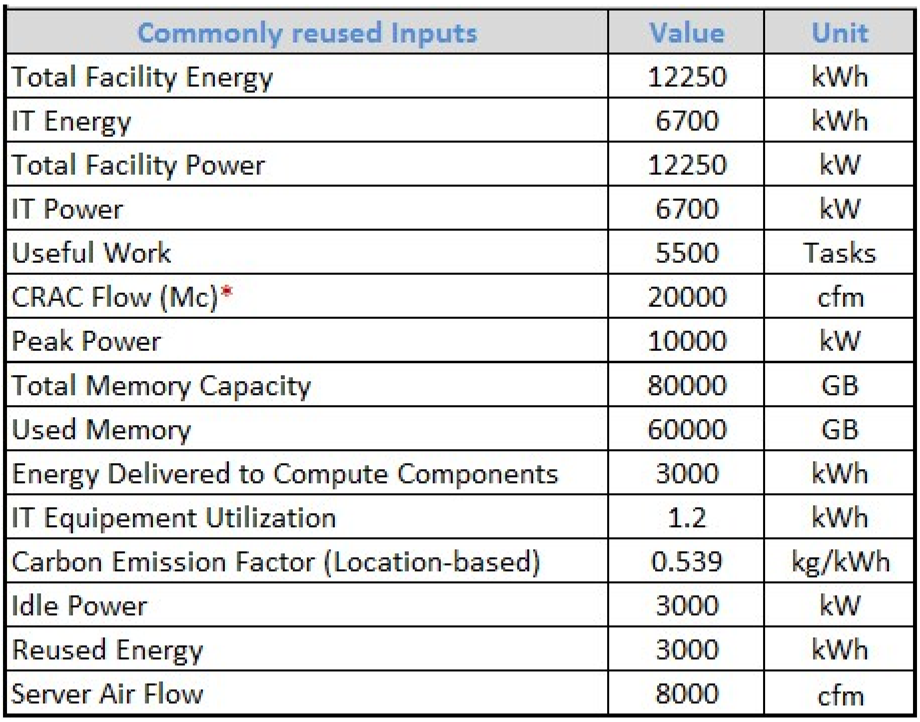


Sustainability Metrics	Unit	Objective	Optimal Value
Power Usage Effectiveness(PUE)*	Ratio	Minimize	1
Data Center infrastructure Efficiency (DCiE)	Percentage	Maximize	1
Carbon Usage Effectiveness (CUE)	KgCO2/kWh	Minimize	0
Data Center Utilization (UDC)	Percentage	Maximize	1
Data Center Performance Efficiency (DCPE)	UW/Power	Maximize	∞
Data Center Energy Productivity (DCeP)	Tasks/ kWhr	Maximize	∞
Energy Reuse Effectiveness (ERE)	Percentage	Minimize	0
Energy Reuse Factor (ERF)	Percentage	Maximize	1
Data Center Productivity (DCP)	Useful work/Watt	Maximize	∞
Idle-to-peak Power Ratio (IPR)	Ratio	Minimize	0
Dynamic Range (DR of Performance)	Ratio	Maximize	1
IT-Power Usage Effectiveness (ITUE)*	Ratio	Minimize	1
Balance Ratio (Air)	Ratio	Maximize	1
Compute Power Efficiency (CPE)	Percentage	Maximize	1
Total-Power Usage Effectiveness (TUE)	Ratio	Minimize	1
Storage Usage	Percentage	Maximize	1

Figure 6.9: The list of sustainability metrics of the sustainable DC simulator

Experiments of the Sustainable Data Center Simulator

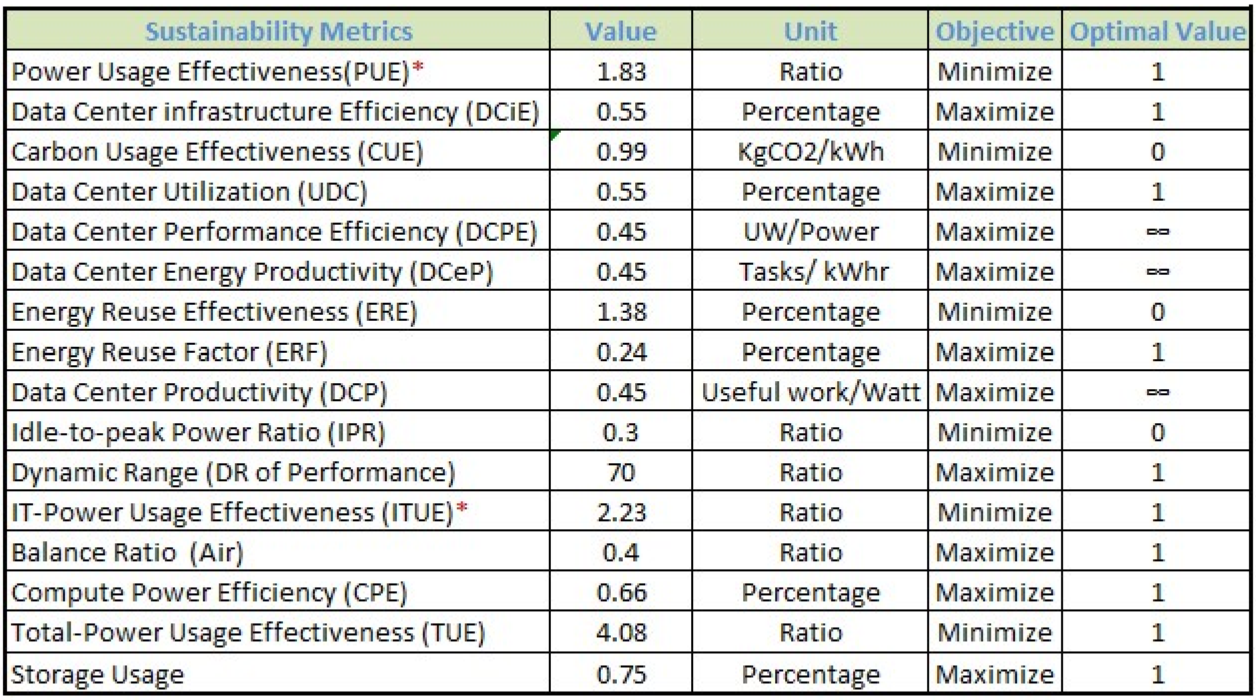
DC operators don't publish the characteristics of DCs, which makes it difficult to find a realistic sample of input parameters that are taken from an actual DC to be used in the experiments. Therefore, we assumed a hypothetical DC with the following input parameters to be simulated using the sustainable DC simulator in the first experiment. Figure 6.10 shows the name, value and unit of each input.



Commonly reused Inputs	Value	Unit
Total Facility Energy	12250	kWh
IT Energy	6700	kWh
Total Facility Power	12250	kW
IT Power	6700	kW
Useful Work	5500	Tasks
CRAC Flow (Mc)*	20000	cfm
Peak Power	10000	kW
Total Memory Capacity	80000	GB
Used Memory	60000	GB
Energy Delivered to Compute Components	3000	kWh
IT Equipement Utilization	1.2	kWh
Carbon Emission Factor (Location-based)	0.539	kg/kWh
Idle Power	3000	kW
Reused Energy	3000	kWh
Server Air Flow	8000	cfm

Figure 6.10: A screenshot of the inputs of the first experiment

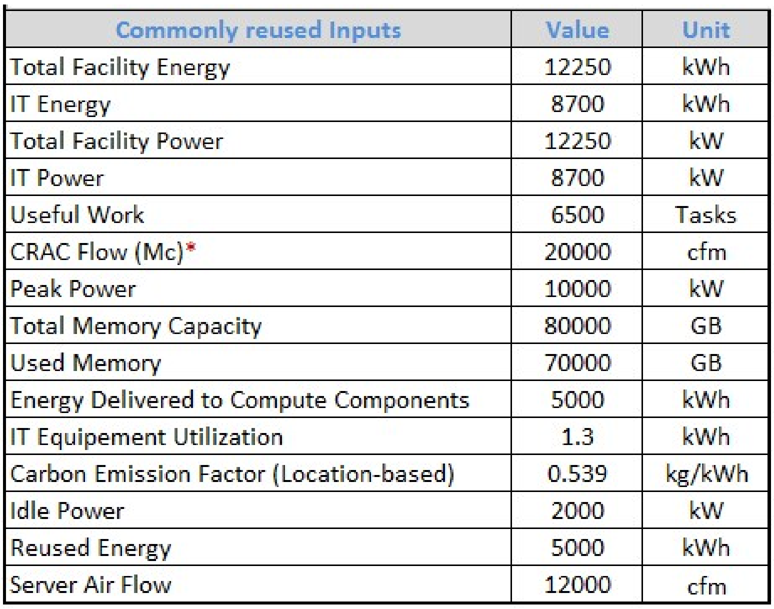
By the time all the inputs are provided in the DC simulator, the sustainability metrics are evaluated. Figure 6.11 shows the list of evaluated metrics, enriched with the optimal value and the intended objective of each metric [VSR+17]. Moreover, the simulator shows that the optimal value of the metrics Data Center Performance Efficiency (DCPE), Data Center Energy Productivity (DCeP), and DCP is infinity, which means that DC operator needs to increase the values of these metrics as much as possible to achieve more sustainability. The mathematical equations that are required for evaluating the metrics are taken from [VSR+17].



Sustainability Metrics	Value	Unit	Objective	Optimal Value
Power Usage Effectiveness(PUE)*	1.83	Ratio	Minimize	1
Data Center infrastructure Efficiency (DCiE)	0.55	Percentage	Maximize	1
Carbon Usage Effectiveness (CUE)	0.99	KgCO2/kWh	Minimize	0
Data Center Utilization (UDC)	0.55	Percentage	Maximize	1
Data Center Performance Efficiency (DCPE)	0.45	UW/Power	Maximize	∞
Data Center Energy Productivity (DCeP)	0.45	Tasks/ kWhr	Maximize	∞
Energy Reuse Effectiveness (ERE)	1.38	Percentage	Minimize	0
Energy Reuse Factor (ERF)	0.24	Percentage	Maximize	1
Data Center Productivity (DCP)	0.45	Useful work/Watt	Maximize	∞
Idle-to-peak Power Ratio (IPR)	0.3	Ratio	Minimize	0
Dynamic Range (DR of Performance)	70	Ratio	Maximize	1
IT-Power Usage Effectiveness (ITUE)*	2.23	Ratio	Minimize	1
Balance Ratio (Air)	0.4	Ratio	Maximize	1
Compute Power Efficiency (CPE)	0.66	Percentage	Maximize	1
Total-Power Usage Effectiveness (TUE)	4.08	Ratio	Minimize	1
Storage Usage	0.75	Percentage	Maximize	1

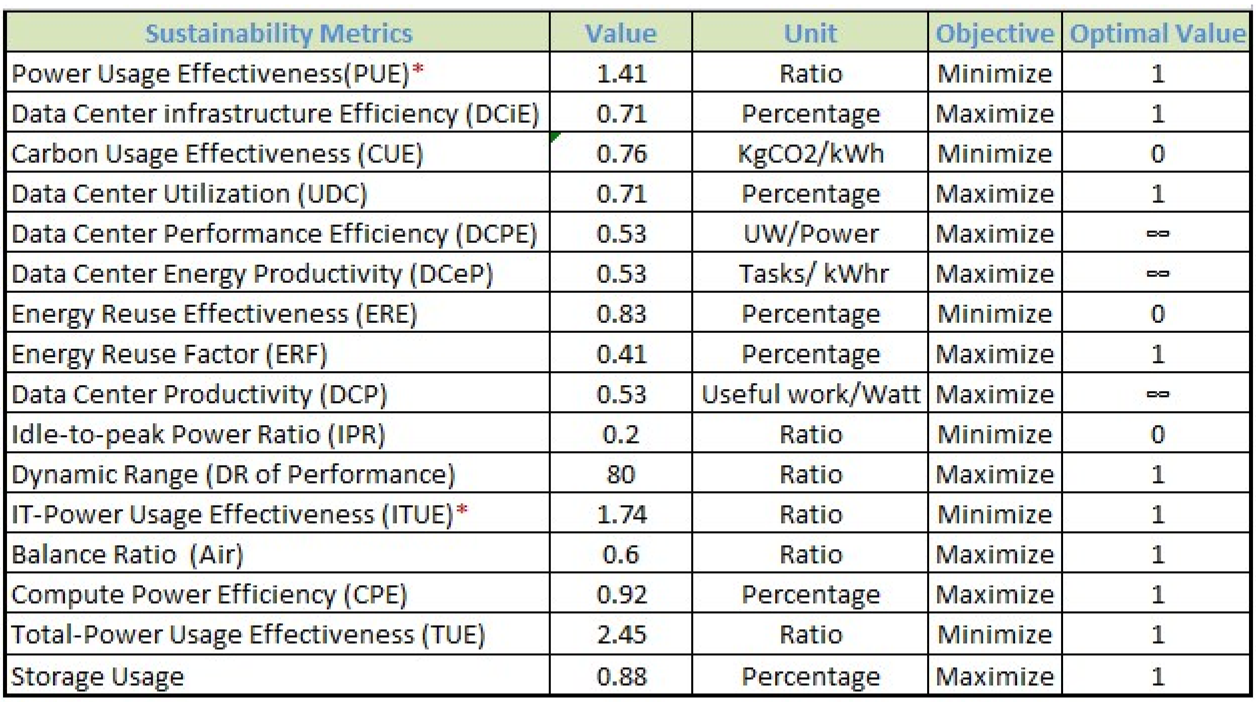
Figure 6.11: A screenshot of the evaluated metrics of the first experiment

In the second experiment, we use the same hypothetical DC inputs but with some changes in some input values such as IT energy, IT power, useful work, used memory, idle power, reused energy, and server airflow. This experiment is done to check how these changes in the workload will impact the sustainability of a DC, and how can the sustainable DC simulator reflect this change through evaluating the different metrics. Figure 6.12 and 6.13 show the inputs and the evaluated metrics.



Commonly reused Inputs	Value	Unit
Total Facility Energy	12250	kWh
IT Energy	8700	kWh
Total Facility Power	12250	kW
IT Power	8700	kW
Useful Work	6500	Tasks
CRAC Flow (Mc)*	20000	cfm
Peak Power	10000	kW
Total Memory Capacity	80000	GB
Used Memory	70000	GB
Energy Delivered to Compute Components	5000	kWh
IT Equipement Utilization	1.3	kWh
Carbon Emission Factor (Location-based)	0.539	kg/kWh
Idle Power	2000	kW
Reused Energy	5000	kWh
Server Air Flow	12000	cfm

Figure 6.12: A screenshot of the inputs of the second experiment



Sustainability Metrics	Value	Unit	Objective	Optimal Value
Power Usage Effectiveness(PUE)*	1.41	Ratio	Minimize	1
Data Center infrastructure Efficiency (DCiE)	0.71	Percentage	Maximize	1
Carbon Usage Effectiveness (CUE)	0.76	KgCO2/kWh	Minimize	0
Data Center Utilization (UDC)	0.71	Percentage	Maximize	1
Data Center Performance Efficiency (DCPE)	0.53	UW/Power	Maximize	∞
Data Center Energy Productivity (DCeP)	0.53	Tasks/ kWhr	Maximize	∞
Energy Reuse Effectiveness (ERE)	0.83	Percentage	Minimize	0
Energy Reuse Factor (ERF)	0.41	Percentage	Maximize	1
Data Center Productivity (DCP)	0.53	Useful work/Watt	Maximize	∞
Idle-to-peak Power Ratio (IPR)	0.2	Ratio	Minimize	0
Dynamic Range (DR of Performance)	80	Ratio	Maximize	1
IT-Power Usage Effectiveness (ITUE)*	1.74	Ratio	Minimize	1
Balance Ratio (Air)	0.6	Ratio	Maximize	1
Compute Power Efficiency (CPE)	0.92	Percentage	Maximize	1
Total-Power Usage Effectiveness (TUE)	2.45	Ratio	Minimize	1
Storage Usage	0.88	Percentage	Maximize	1

Figure 6.13: A screenshot of the evaluated metrics of the second experiment

The following two tables summarize the two experiments. Table 6.2 shows the name of inputs, values of the first experiment, values of the second experiment, and the difference between both. Similarly, Table 6.3 shows the values of the evaluated metrics after the two experiments. Moreover, the difference between the values of both experiments is provided in a separate column.

Table 6.2: The values of inputs of both experiments

Input Name	Experiment 1	Experiment 2	Difference
Total Facility Energy	12250	12250	0
IT Energy	6700	8700	+2000
Total Facility Power	12250	12250	0
IT Power	6700	8700	+2000
Useful Work	5500	6500	+1000
CRAC Flow (Mc)	20000	20000	0
Peak Power	10000	10000	0
Total Memory Capacity	80000	80000	0
Used Memory	60000	70000	+10000
Delivered Energy	3000	5000	+2000
IT Equipment Utilization	1.2	1.3	+0.1
Carbon Emission Factor	0.539	0.539	0
Idle Power	3000	2000	-1000
Reused Energy	3000	5000	+2000
Server Air Flow	8000	12000	+4000

Table 6.3: The values of evaluated metrics of both experiments

Metric Name	Experiment 1	Experiment 2	Difference
Power Usage Effectiveness PUE	1.83	1.41	-0.42
Data Center infrastructure Eff. DCiE	0.55	0.71	+0.16
Carbon Usage Effectiveness CUE	0.99	0.76	-0.23
Data Center Utilization UDC	0.55	0.71	+0.16
Data Center Performance Eff. DCPE	0.45	0.53	+0.08
Data Center Energy Productivity DCeP	0.45	0.53	+0.08
Energy Reuse Effectiveness ERE	1.38	0.83	-0.55
Energy Reuse Factor ERF	0.24	0.41	+0.17
Data Center Productivity	0.45	0.53	+0.08
Idle-to-peak Power Ratio IPR	0.3	0.2	-0.1
Dynamic Range	70	80	+10
IT-Power Usage Effectiveness ITUE	2.23	1.74	-0.49
Balance Ratio (Air)	0.4	0.6	+0.2
Compute Power Efficiency (CPE)	0.66	0.92	+0.26
Total-Power Usage Effectiveness (TUE)	4.08	2.45	-1.63
Storage Usage	0.75	0.88	+0.13

7 Discussion

7.1 Analysis Results

In the metrics analysis chapter, a tree graph is created for each category of DC sustainability metrics to illustrate the connection between the input parameters and the metrics, each metric is connected to the inputs from which it is calculated. The inputs that are reused more than once are called as commonly reused inputs, while the metrics that are reused one time or more are called as commonly reused metrics. On the graphs, the commonly reused inputs are highlighted in yellow and the commonly reused metrics are highlighted in blue to show how significant they are. Table 5.1 is created to summarize the Commonly reused inputs and the frequency of repetition. Similarly, Table 5.2 is created to list the commonly reused metrics and the frequency. Afterwards, a rooted tree graph is created to show the commonly reused input or metric as a parent node, and the siblings are the metrics that use this input to be calculated.

RQ 1: What are the commonly reused inputs and metrics in a DC?

Finally, the commonly reused inputs and metrics are analysed to identify the maximum number of sustainability metrics that can be evaluated from the minimum number of commonly reused inputs, and this is the answer of the first research question. Table 7.1 shows the final list of 15 commonly reused inputs which can evaluate 16 sustainability metrics.

Table 7.1: The list of 15 commonly reused Inputs that evaluate 16 metrics

Commonly Reused Inputs	Evaluated Metrics
Total Facility Energy	Power Usage Effectiveness (PUE)
IT Energy	Data Center Infrastructure Efficiency (DCiE)
Total Facility Power	Carbon Usage Effectiveness (CUE)
IT Power	Storage Usage
Useful Work	Data Center Performance Efficiency (DCPE)
CRAC Flow (Mc)	Data Center Energy Productivity (DCeP)
Peak Power	Energy Reuse Effectiveness (ERE)
Total Memory Capacity	Energy Reuse Factor (ERF)
Used Memory	Data Center Productivity (DCP)
Energy Delivered to Components	Idle-to-peak Power Ratio (IPR)
IT Equipment Utilization	Dynamic Range (DR)
Carbon Emission Factor	IT-Power Usage Effectiveness (ITUE)
Idle Power	Balance Ratio
Reused Energy	Total-Power Usage Effectiveness (TUE)

Continued on next page

Table 7.1 – Continued from previous page

Commonly Reused Inputs	Evaluated Metrics
Server Air Flow	Compute Power Efficiency (CPE) Data Center Utilization (UDC)

Table 7.1 lists PUE as one of the evaluated metrics, PUE is the most popular energy efficiency metric, it is used by a large number of other metrics either directly or as a derivation [VSR+17]. A PUE of 1.0 means 100% of the power brought to the DC goes to IT equipment and none to cooling, lighting, or other non-IT loads [TVCA10].Additionally, CUE is listed also to be evaluated, it is emerging as an extremely important factor in the design, location, and operation of DC facilities.

7.2 Simulation Results

The most relevant DC simulation platforms are investigated to check if it is possible to use any of them to employ our approach of evaluating 16 metrics by using 15 commonly reused input. But the comparison in Table 6.1 shows that, there is no simulator that can fulfill all the requirements.

RQ 2: How can we use the commonly reused inputs and metrics to simulate a sustainable DC?

Since there is no single simulator that fulfill our requirements, we designed the sustainable DC simulator. IT uses the 15 commonly reused inputs to evaluate 16 of the sustainability metrics. All the required calculations are incorporated in the excel-based tool to facilitate the simulation and obtain the results. The evaluated metrics cover 5 different metrics categories, the evaluated metrics include metrics that contribute to reflect the sustainability of a DC, such as PUE, CUE, DCPE, Data Center Utilization DCU, DCP, Energy Reuse Effectiveness (ERE), Total Power Usage Effectiveness TUE, Balance Raio BR, and Storage Usage of a DC. Moreover, the objective and the optimal value for each metric is provided as a reference to guide the DC operators.

In the last chapter, Table 6.1 and 6.2 list all the inputs and results of the two experiments. Figure 7.1 shows a visual representation of the results of 5 metrics. In the first simulation experiment, the input value of IT energy was 6700 kWh, and the corresponding results of the evaluated metrics are in red. In the second simulation, the input value of IT energy increased to be 8700 kWh, and with this change in one commonly reused input, we can clearly see the influence on all the metrics values, the 5 metrics are optimized towards the objectives.

The bar chart in Figure 7.1 shows that, the value of PUE is decreased just by changing the value of IT energy. The metric DCiE, which is the inverse of PUE [Elec], is affected accordingly. But the objective of DCiE is to be maximized, and that happened also by increasing the value of IT energy. DCiE metric is also useful, DCiE value of 0.71 indicated that the IT equipment consumes 71% of the power in this DC.

We assume that the location of the DC is Germany therefore, it's carbon emission factor is 0.539 kg/kWh because this factor is location-based [Elea]. Carbon Usage Effectiveness is considered the most important metric related to environmental sustainability of a DC, defined by The Green Grid [APPT10], it considers the carbon footprint of a given DC, measuring the greenhouse gas emissions in relation to IT energy consumption.

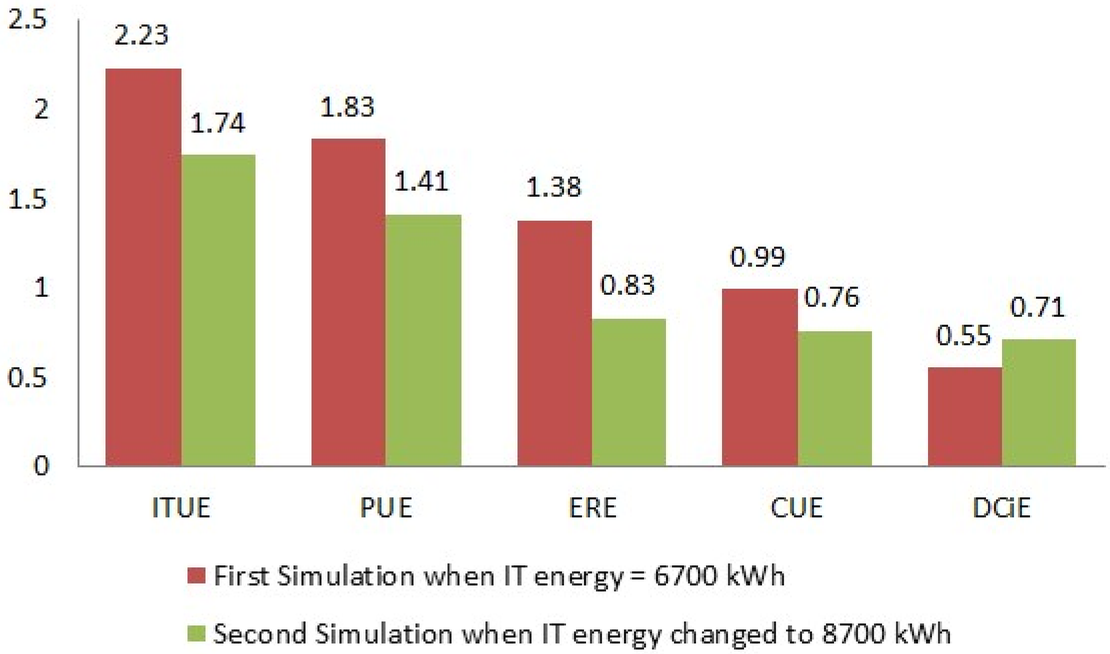


Figure 7.1: Visualizing metrics that are related to energy efficiency, greenness, and reused energy of the DC

Figure 7.1 shows that CUE is also affected by the change in the input value of IT energy. CUE has an ideal value of 0, which indicates that no carbon use is associated with the DC's operations. From the experiments results, CUE is decreased in the simulator, but it needs more optimization. Evaluation the CUE metric helps the DC operators to determine the environmental impact of their DCs' CO2 emissions. The metrics CUE and PUE enable DC operators to quickly assess the relative sustainability of their DCs, compare the results, and determine if any sustainable energy improvements need to be made [APPT10].

Increasing the value of one or two commonly reused inputs, can provide DC operators with precious insights, because it reflects a direct positive change in 5 sustainability metrics. The metrics ERE and ITPUE are optimized as well, not only with the change in the input value of IT energy, but also with the change in the input value of reused energy, since this commonly reused input is also used besides IT energy value to calculate the two metrics. For ERE, the range is 0 to infinity, it allows values less than 1.0. An ERE of 0 means that 100% of the energy brought into the DC is reused elsewhere, which is the optimal in this case [TVCA10]. Figure 7.1 shows that ERE is optimized by changing the input values of reused energy and IT energy. Nevertheless, the value of 0.83 for ERE is not optimal and DC operators in this case, need to increase the amount of reused energy in order to achieve more sustainability in their DCs.

Figure 7.2 visualizes storage usage, which is one of the storage metrics that measures the percentage of storage used relative to the overall storage capacity within a DC, the change in the input value of used memory directly influence this metric. the bar chart shows the optimized value of storage usage which is 88%. But, since of 70% is believed to be ideal percentage of storage usage [VSR+17], the DC operator in this case has to adjust it by either increasing the storage capabilities, or by decreasing the used memory.

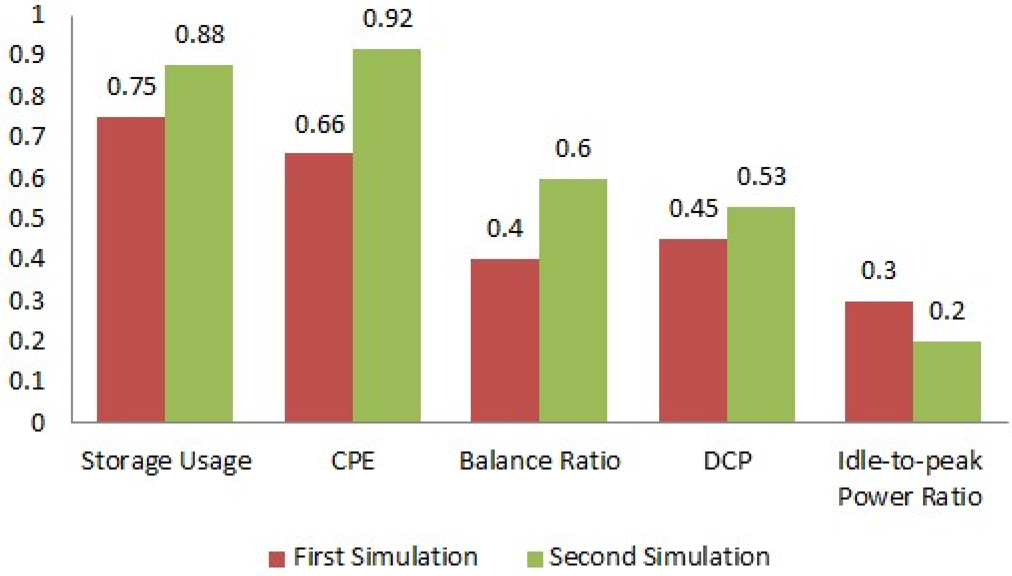


Figure 7.2: Visualising metrics that are related to storage, air management, and productivity of the DC

The balance ratio of 0.5 indicates that the servers are not getting enough air [TS10]. But in case of the simulated DC, balance ratio increased from 40% in the first experiment to 60% in the second experiment. This increase resulted from the change in the input value of server air flow, while the CRAC which is the total air supplied by the cooling units in a DC doesn't change in both experiments. Figure 7.2 shows that, DC productivity increased by increasing the input value of useful work. Idle-to-peak power ration is optimized to be 0.2 instead of 0.3 in the first experiment, this is achieved by decreasing the input value of idle power.

From the results of the two simulation experiments, we can conclude that the sustainable DC simulator can help DC operators in evaluating 16 sustainability metrics by using 15 commonly reused inputs, which validates our main approach of evaluating the maximum number of sustainability metrics from the minimum number of commonly reused inputs. The DC operators can use the simulator to monitor the change in workload, because by changing few number of commonly reused inputs, they can evaluate a lot of sustainability metrics. Additionally, the evaluated metrics include some important metrics such as PUE, DCiE, DCeP, DCPE, ITPUE, and TUE which reflects the energy efficiency of a DC, and the metrics CUE, ERE, and ERF which reflects the greenness of a DC, and the metric balance ratio which partially helps in improving the air management of a DC, and the metrics DCeP and DCP reflect the productivity of a DC.

RQ 3: How can this DC simulator help DC operators to achieve sustainability?

The evaluated metrics give insights and indications to DC operators to improve and optimize the performance and energy efficiency of DCs. Especially, the CUE metric which indicates the carbon dioxide usage effectiveness. Consequently, by improving the energy efficiency and decreasing the carbon footprint, DC operators can achieve higher level of sustainability in DCs. Hence, moving step forward in solving the DC challenges that were discussed in the problem statement section, which are related to the enormous amounts of electrical power consumption and the resulting carbon emissions.

8 Conclusion and Future Work

Data centers are heavy power consumers, they require a tremendous amount of energy, which can be harmful to the environment due to the resulting carbon dioxide emissions. Moreover, the existing DC simulators focus more on the resource management and deployment of applications as we discussed in the problem statement. Therefore, we built a sustainable DC, that can better monitor the DC power consumption hence, reduce the emissions, this research focuses on designing a sustainable DC simulator, which can help DC operators to determine the impact of changing the workload on the sustainability of a DC. The workload of a DC is represented by the input parameters of the simulator, while the DC sustainability can be evaluated by the sustainability metrics. We provided a profound analysis of the input parameters and sustainability metrics in chapter 5, to identify the commonly reused inputs and metrics. The analysis led to a set of commonly reused inputs. We used the minimum number of commonly reused inputs to evaluate the maximum number of sustainability metrics. We were able to evaluate 16 metrics from 15 commonly reused inputs, and this is the main idea of our approach to build the sustainable DC simulator. Afterwards, we designed the sustainable DC, and 2 experiments were applied to demonstrate how can DC operators achieve sustainability in DCs using the proposed simulator, and we addressed determining the impact of changes in workload and how the consequences affect the sustainability of a DC.

The main limitation of this research, was to find a realistic sample of DC inputs to use it in the experiments. DC operators don't publish data about their DCs, and the lack of reliable data limits the experiments to the assumed sample of input parameters, which might have some inconsistencies.

The future work can be done to include the other metrics categories into the simulator. Metrics categories such as cooling, network, security, and financial are not yet represented in the simulator because of the selection that we made to limit the number of input parameters. Therefore, the excluded metrics categories require additional work to be considered so that, the sustainability of a DC can be evaluated in a holistic manner. Cooling equipment consumes large amounts of power and further research in the area of CFD would help in optimizing cooling, which in turn can be beneficial for DC operators in order to achieve more power-efficient DCs. Another interesting aspect to address, is the visualisation of results in the simulator.

Bibliography

[AAFP12] V. Avelar, D. Azevedo, A. French, E. N. Power. “PUE: a comprehensive examination of the metric”. In: *White paper* 49 (2012) (cit. on p. 23).

[ana07] F. T. G. S. U. U. I. M. analysis. “Advanced Micro Devices”. In: (2007) (cit. on p. 24).

[APPT10] D. Azevedo, M. Patterson, J. Pouchet, R. Tipley. “Carbon usage effectiveness (CUE): a green grid data center sustainability metric”. In: *The green grid* (2010) (cit. on pp. 16, 82, 83).

[Bel05] M. Bell. “Use Best Practices to Design Data Center Facilities”. In: (Jan. 2005) (cit. on pp. 21, 25, 28).

[BM+76] J. A. Bondy, U. S. R. Murty, et al. *Graph theory with applications*. Vol. 290. Macmillan London, 1976 (cit. on p. 31).

[BM10] C. L. Belady, C. G. Malone. “Metrics and an infrastructure model to evaluate data center efficiency”. In: *ASME 2007 InterPACK Conference collocated with the ASME/JSME 2007 Thermal Engineering Heat Transfer Summer Conference*. American Society of Mechanical Engineers Digital Collection. 2010, pp. 751–755 (cit. on p. 24).

[Bou11] D. Bouley. “Estimating a data center’s electrical carbon footprint”. In: *Schneider Electric White Paper Library* (2011) (cit. on pp. 14, 22, 23, 33).

[BP10] R. Baskerville, J. Pries-Heje. “Explanatory Design Theory”. In: *Business & Information Systems Engineering* 2.5 (Oct. 2010), pp. 271–282. ISSN: 1867-0202. DOI: 10.1007/s12599-010-0118-4. URL: https://doi.org/10.1007/s12599-010-0118-4 (cit. on p. 29).

[BPSS08] C. Bash, C. Patel, A. Shah, R. Sharma. “The sustainable information technology ecosystem”. In: June 2008, pp. 1126–1131. ISBN: 978-1-4244-1700-1. DOI: 10.1109/ITHERM.2008.4544387 (cit. on p. 16).

[Bro+08] R. Brown et al. “Report to congress on server and data center energy efficiency: Public law 109-431”. In: (2008) (cit. on p. 16).

[CdL+18] L. Chiaraviglio, F. d’Andreagiovanni, R. Lancellotti, M. Shojafar, N. Blefari-Melazzi, C. Canali. “An approach to balance maintenance costs and electricity consumption in cloud data centers”. In: *IEEE Transactions on Sustainable Computing* 3.4 (2018), pp. 274–288 (cit. on p. 20).

[CRDB09] R. N. Calheiros, R. Ranjan, C. A. De Rose, R. Buyya. “Cloudsim: A novel framework for modeling and simulation of cloud computing infrastructures and services”. In: *arXiv preprint arXiv:0903.2525* (2009) (cit. on p. 65).

[Elea] S. Electric. *Schneider Electric, DATA CENTER CARBON FOOTPRINT COMPARISON CALCULATOR*. URL: `https://www.schneider-electric.com/en/work/solutions/system/s1/data-center-and-network-systems/trade-off-tools/data-center-carbon-footprint-comparison-calculator/`. (accessed: 20.09.2019) (cit. on p. 82).

[Eleb] S. Electric. *Schneider Electric, Data Center Efficiency Calculator – A Tool for Modeling Your Current and/or Future PUE*. URL: `https://blog.se.com/datacenter/2012/04/30/data-center-efficiency-calculator-a-tool-for-modeling-your-current-andor-future-pue/`. (accessed: 10.09.2019) (cit. on p. 70).

[Elec] S. Electric. *Schneider Electric, DATA CENTER EFFICIENCY PUE CALCULATOR*. URL: `https://www.schneider-electric.com/en/work/solutions/system/s1/data-center-and-network-systems/trade-off-tools/data-center-efficiency-and-pue-calculator/`. (accessed: 15.09.2019) (cit. on pp. 34, 70–72, 82).

[Eled] S. Electric. *Schneider Electric Data Center TradeOff Tools*. URL: `https://www.schneider-electric.com/en/work/solutions/for-business/data-centers-and-networks/trade-off-tools`. (accessed: 16.09.2019) (cit. on p. 70).

[Enea] T. C. O. E. for Energy Efficiency in Data Centers. *Data Center Profiler (DC Pro) Tools*. URL: `https://datacenters.lbl.gov/dcpro`. (accessed: 05.09.2019) (cit. on pp. 69, 70).

[Eneb] T. C. O. E. for Energy Efficiency in Data Centers. *The CENTER OF EXPERTISE for Energy Efficiency in Data Centers*. URL: `https://datacenters.lbl.gov/tools`. (accessed: 09.09.2019) (cit. on p. 69).

[GCS+10] D. Gmach, Y. Chen, A. Shah, J. Rolia, C. Bash, T. Christian, R. Sharma. “Profiling Sustainability of Data Centers”. In: June 2010, pp. 1–6. DOI: `10.1109/ISSST.2010.5507750` (cit. on p. 26).

[Gel] J. D. Gelas. *Green Sustainable Data Centres Measurement and Control*. URL: `https://www.ou.nl/documents/380238/382808/GSDC_05_Measurement_and_control.pdf/e58f5aa9-3cad-dccd-b15e-83cbb778f3c9`. (accessed: 05.09.2019) (cit. on p. 15).

[GGB+11] S. K. Gupta, R. R. Gilbert, A. Banerjee, Z. Abbasi, T. Mukherjee, G. Varsamopoulos. “Gdcsim: A tool for analyzing green data center design and resource management techniques”. In: *2011 International Green Computing Conference and Workshops*. IEEE. 2011, pp. 1–8 (cit. on pp. 27, 67, 68).

[GKL+13] Í. Goiri, W. Katsak, K. Le, T. D. Nguyen, R. Bianchini. “Parasol and greenswitch: Managing datacenters powered by renewable energy”. In: *ACM SIGPLAN Notices*. Vol. 48. 4. ACM. 2013, pp. 51–64 (cit. on pp. 66, 67).

[GY05] J. L. Gross, J. Yellen. *Graph theory and its applications*. CRC press, 2005 (cit. on pp. 29, 51, 58).

[KBSW10] J. Koomey, S. Berard, M. Sanchez, H. Wong. “Implications of historical trends in the electrical efficiency of computing”. In: *IEEE Annals of the History of Computing* 33.3 (2010), pp. 46–54 (cit. on p. 13).

[Koo+11] J. Koomey et al. “Growth in data center electricity use 2005 to 2010”. In: *A report by Analytical Press, completed at the request of The New York Times* 9 (2011), p. 161 (cit. on p. 26).

[LSN+09] S.-H. Lim, B. Sharma, G. Nam, E.-K. Kim, C. R. Das. "MDCSim: A multi-tier data center simulation, platform." In: *CLUSTER*. Vol. 31. 2009, pp. 1–9 (cit. on p. 63).

[MBA+14] A. W. Malik, K. Bilal, K. Aziz, D. Kliazovich, N. Ghani, S. U. Khan, R. Buyya. "Cloudnetsim++: A toolkit for data center simulations in omnet++". In: *2014 11th Annual High Capacity Optical Networks and Emerging/Enabling Technologies (Photonics for Energy)*. IEEE. 2014, pp. 104–108 (cit. on p. 66).

[MMS+12] B. W. MEng, C. M. M. MASHRAE, A. Shah, B. K. P. C. B. MASHRAE, C. M. FInstR, D. A. M. R. Cert, A. D. B. P. D MCIBSE. "The Environmental Burden of Data Centres–A Screening LCA Methodology". In: (2012) (cit. on pp. 27, 28).

[Mod] Modius. *Smart Data Center Infrastructure Management*. URL: `http://modius.com/smartdatacenter/`. (accessed: 12.09.2019) (cit. on p. 20).

[Mur08] S. Murugesan. "Harnessing green IT: Principles and practices". In: *IT professional* 10.1 (2008), pp. 24–33 (cit. on p. 38).

[NB12] F. Norouzi, M. A. Bauer. "Toward an autonomic energy efficient data center". In: *2012 IEEE International Conference on Green Computing and Communications*. IEEE. 2012, pp. 487–493 (cit. on p. 68).

[New08] L. Newcombe. "Data centre energy efficiency metrics: Existing and proposed metrics to provide effective understanding and reporting of data centre energy". In: *BCS: British Computer Society* (2008) (cit. on pp. 13, 14).

[OPPF10] S. Ostermann, K. Plankensteiner, R. Prodan, T. Fahringer. "GroudSim: an event-based simulation framework for computational grids and clouds". In: *European Conference on Parallel Processing*. Springer. 2010, pp. 305–313 (cit. on p. 66).

[PBS+03] C. Patel, C. Bash, R. Sharma, M. Beitelmal, R. Friedrich. "Smart Cooling of Data Centers". In: Jan. 2003. DOI: `10.1115/IPACK2003-35059` (cit. on p. 23).

[RMM15] B. Ristic, K. Madani, Z. Makuch. "The water footprint of data centers". In: *Sustainability* 7.8 (2015), pp. 11260–11284 (cit. on p. 25).

[Rom19] J. Romm. "Data centers, power, and pollution prevention : design for business and environmental advantage". In: (Sept. 2019) (cit. on pp. 15, 19, 26).

[RR17] H. Ritchie, M. Roser. "CO and Greenhouse Gas Emissions". In: *Our World in Data* (2017) (cit. on p. 22).

[Sima] Simscale. *How to Reduce Data Center Power Consumption with CFD*. URL: `https://www.simscale.com/blog/2018/05/data-center-power-consumption/`. (accessed: 14.09.2019) (cit. on p. 19).

[Simb] Simscale. *Simscale Documentation*. URL: `https://www.simscale.com/docs/`. (accessed: 01.05.2019) (cit. on p. 68).

[TKBL12] M. Tighe, G. Keller, M. Bauer, H. Lutfiyya. "DCSim: A data centre simulation tool for evaluating dynamic virtualized resource management". In: *2012 8th international conference on network and service management (cnsm) and 2012 workshop on systems virtualiztion management (svm)*. IEEE. 2012, pp. 385–392 (cit. on pp. 27, 63, 64).

[TS10] R. Tozer, M. Salim. "Data center air management metrics-practical approach". In: *2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*. IEEE. 2010, pp. 1–8 (cit. on p. 84).

[TVCA10] B. Tschudi, O. Vangeet, J. Cooley, D. Azevedo. “ERE: A metric for measuring the benefit of reuse energy from a data center”. In: *White Paper* 29 (2010) (cit. on pp. 82, 83).

[UA12] M. Uddin, A. Abdul Rahman. “Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics”. In: *Renewable and Sustainable Energy Reviews* 16 (Aug. 2012), pp. 4078–4094. DOI: `10.1016/j.rser.2012.03.014` (cit. on pp. 16, 19).

[VDT18] T. Van Damme, C. De Persis, P. Tesi. “Optimized thermal-aware job scheduling and control of data centers”. In: *IEEE Transactions on Control Systems Technology* 27.2 (2018), pp. 760–771 (cit. on p. 26).

[VKSK17] V. Vijayan, B. Kuriakose, S. Sooraj, N. A. Kumar. “Components of Data Center Environment”. In: *International Journal of Pure and Applied Mathematics* 114.12 (2017), pp. 71–78 (cit. on p. 21).

[VLL+14] W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet, P. Demeester. “Trends in worldwide ICT electricity consumption from 2007 to 2012”. In: *Computer Communications* 50 (2014), pp. 64–76 (cit. on p. 20).

[VSR+17] D.-D. R. Vemula, B. Setz, G. Rao, G. R. Gangadharan, M. Aiello. “Metrics for Sustainable Data Centers”. In: *IEEE Transactions on Sustainable Computing* PP (May 2017), pp. 1–1. DOI: `10.1109/TSUSC.2017.2701883` (cit. on pp. 4, 16, 23, 25, 26, 29, 31, 33, 41, 44, 49, 77, 82, 84).

[VSR+18] D.-D. R. Vemula, B. Setz, G. Rao, G. R. Gangadharan, M. Aiello. “Best Practices for Sustainable Datacenters”. In: *IT Professional* 20 (Sept. 2018), pp. 57–67. DOI: `10.1109/MITP.2018.053891338` (cit. on pp. 26, 28).

[WASM14] B. Whitehead, D. Andrews, A. Shah, G. Maidment. “Assessing the environmental impact of data centres part 1: Background, energy use and metrics”. In: *Building and Environment* 82 (Dec. 2014), pp. 151–159. DOI: `10.1016/j.buildenv.2014.08.021` (cit. on p. 26).

[Wel] Welcomeitalia. *Data center.* URL: `https://corporate.welcomeitalia.it/en/data-center/`. (accessed: 13.09.2019) (cit. on p. 21).

All links were last followed on September 30, 2019.

A The List of Input parameters of a DC

Table A.1: The list of inputs.

Abbreviation	**Unit**	**Description**
EDCi	kWh	The DC energy consumption
EPi	kWh	The planned energy
KAPC	Factor	The adjustment factor for normalizing EDCi and EPi
FE	Ratio	Facility efficiency
AE	Ratio	Asset efficiency
IT Utilization	Ratio	The average CPU utilization
IT Energy Efficiency	Ratio	Efficiency of IT components
Facility Energy Eff.	Ratio	Facility energy efficiency is the ratio of IT load to total power
Facility Utilization	Ratio	The ratio of the Aatual IT load over the DC capacity
ITEU	Percentage	IT equipment utilization
Total Facility Power	kWh	total facility power
IT Equipment Power	kWh	The total power of IT components
PUE	Percentage	Power usage effectiveness
EDCReal i	kWh	Energy consumption after changing operational modes
EDCBaseline i	kWh	The energy consumption before changing operational modes
KDCA	Factor	The adjustment factor for normalizing the EDCReal i and EDCBaseline i energy curves
ScE	Percentage	Server compute efficiency
k	Number	The total number of servers
W	Percentage	Useful work performed
Total Facility Energy	kWh/Houre	Energy consumption by DC
Power for DC lighting	kW	Power consumption for DC lighting
Total DC space	square foot	Total area of the DC
Variable energy	kWh/Houre	Variable percentage of DC energy consumption
Fixed Energy	kWh/Houre	Fixed percentage of DC energy consumption
Total number of servers deployed	Number	Total number of servers deployed

Continued on next page

Table A.1 – Continued from previous page

Abbreviation	Unit	Description
M	Number	The minimum number of servers required to compute peak load
Live Servers	Number	Number of servers which are live executing some applications
Total Memory deployed	GB	Total memory deployed
Last 90 days accessed data	GB	Amount of frequently accessed data within the last 90 days
WPE	kWh/Houre	Energy efficiency of a specific workload
sPUE	Percentage	The overhead for operating a given system in a certain DC
Cost ei	Euro/kWh	The electricity cost
Eoth DCi	kWh/Houre	The total primary energy produced from other sources
Cost othi	Euro/kWh	The electricity cost per kWh from other sources
EDC-	kWh	Energy of DC related to useless workload
Total Hardware load (AC)	kWh	Total hardware load of DC at the plug(AC)
Total compute load(DC)	kWh	Total hardware compute load
Total IT Energy	kWh/Houre	Total energy into IT equipment
Energy of IT Components	kWh/Houre	Total energy delivered to IT components
Count OS	Number	Count of all OS instances
PDC	kWh	The total facilities power at the time of assessment
P inf.	kWh	The power draw of the supporting infrastructure
P IT	kWh	The power consumed by the IT equipment in the rack
IT Equipment Energy	kWh/Houre	The amount of power delivered to run the IT equipment such as servers
mPUE	Ratio	The total facility energy usage divided by the IT energy usage.
m actual	Ratio	The linear relationship of IT power (PIT) to facility power
Energy of a subsection	kWh/Houre	Total energy delivered to a subsection
Energy by IT eq. of subsection	kWh/Houre	Energy used by IT equipment inside that subsection
Cooling hours by Air Economizer	Number	Free cooling hours by air economizer
Heat extracted by Air Conditioners	Joule	Heat extracted by air conditioners
Work Input of cooling System	kWh/Houre	Work Input of cooling System

Continued on next page

Table A.1 – Continued from previous page

Abbreviation	Unit	Description
Average cooling power usage	kWh	Average cooling system power usage
Average cooling load	Ton	Average cooling load in the DC
TC	Ton	The total cooling capacity
Q cooling	Wh	The heat removed by the cooling system
P Cooling	Wh	The electrical power used by the cooling system
Water Economizer Hours	Number	Water economizer hours (full cooling)
CO2 ei (base and current)	Ton	The total CO2 emissions released based on the grid energy consumed by the DC
CO2 othi (base and current)	Ton	The total CO2 emissions released by the energy produced from other sources
CO2 emission of the Total DC energy	Ton	CO2 emission caused by the Total DC energy
Total weight responsibly disposed	Ton	How well EOL equipment is managed through certified entities
Total weight all decommissioned	Ton	The total weight of all equipment that is either reused, recycled, or wasted
Energy Reused	kWh/Houre	Energy reused
Renewable energy	kWh	Renewable energy used in kWh
EEi	Ratio	The ratio of performance to power consumed
EERefi	Ratio	The ratio of Performance to power consumption when executing the benchmark
Total Material	Ton	All the material (recycled + reclaimed + repurposed)
Total Inbound Material	Ton	Total inbound material
Total Outbound products	Number	Total outbound products or services
En	kWh/Houre	The energy consumption associated with indirect water usage and direct water usage
Ed	kWh/Houre	The energy consumption associated with indirect water usage and direct water usage
ECER	Ratio	Electricity carbon emission rate
Annual Water Usage	Liters	Annual water usage
ASEWU	Number	The annual source energy water usage
ASWU	Number	Annual site Wwter usage
EWIF	Factor	Energy water intensity factor
Used CPU	GHz	Used CPU
Allocated CPU	BC	Allocated CPU
IT-PEW	Wh	Measures the electricity consumption in terms of productivity
SI-EER	Ratio	Site infrastructure energy efficiency Ratio

Continued on next page

Table A.1 – Continued from previous page

Abbreviation	Unit	Description
Peak Power	Wh	The power consumption at 100% utilization
Idle Power	Wh	The power consumption at 0% utilization
Area actual	Square Feet	The area under the server's actual energy proportionality curves
Area linear	Square Feet	The area under the server's linear energy proportionality curves
Actual Power	BC	The actual power
Ideal Power	BC	The ideal power
UPS Output Power	kWh	UPS output power
UPS Input Power	kWh	UPS input power
Apparent Power	kWh	The apparent power
Total fan power	kWh	Total fan power includes supply and return
Total fan airflow	kWh	Total fan air flow includes supply and exhaust
Total heat load	Wh	Total heat load
TCPU max	Celsius	It is the maximum temperature reached by the CPUs in the node-group at a given time-stamp
TCPU min	Celsius	It is the minimum temperature reached by the CPUs in the node-group at a given time-stamp
TCPU maxref	Celsius	It is the reference value for the maximum acceptable temperature
Actual vapor density	g/m3	Actual vapor density
Saturation vapor density	g/m3	Saturation vapor density
Total Over-Temp	Celsius	Total Over-Temperature
Max Allowable Over-Temp	Celsius	Max allowable over-temperature
Total Under-Temp	Celsius	Total under-temperature
Max Allowable Under-Temp	Celsius	Max allowable under-temperature
Power Consumed by Network Equipment	Wh	Power consumed by network equipment
Effective Network Throughput Capacity	Kbps	Effective network throughput capacity
Outbound Bits	Number	Outbound bits
Sj	Number	It is the sum of servers in each connected component j
Memory Capacity	Gb	Memory capacity of the DC
Used Memory by a server	GB	Used memory by a server or application
Allocated Main Memory	GB	Allocated main memory
I/O operations per second	Number	I/O operations per second
Mega Bytes moved per second	Number	Mega bytes moved per second
Fi	Number	It is the hit count for rule ri

Continued on next page

Table A.1 – Continued from previous page

Abbreviation	Unit	Description
IAS	Number	It is the interface accessibility surface of an interface
Pd	Percentage	It is Probability of detection in test cases
Pfa	Percentage	It is the probability of false alarm
Total number of destinations	Number	Total number of destinations addresses
Ns	Number	It is the number of ports that reply to traffic from a source
No	Number	It is the number of machines that have two-way connection-oriented sessions to the point of origin
Np	Number	It is the number of paths that have physical access to secure parts such as storage drives
Ti	Date	It is the date on which the ith vulnerability is discovered
MTBF	Time	Represents the average time between consecutive failures or outages of a DC
Total surviving hours	Number	Total surviving hours
Number of failures or outages	Number	Number of failures or outages
MTTR	Time	Gives the expected time to failure for a non-repairable component
MTTF	Time	Represents the average time it takes to repair components or solve problems
Net profit	Euro	Net profit
cost of investment	Euro	cost of investment

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature