UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ELETRÔNICA E SISTEMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

KÁDNA MARIA ALVES CAMBOIM VALE

ANALYSIS OF DEPENDABILITY AND SUSTAINABILITY REQUIREMENTS TO SUPPORT THE DEPLOYMENT OF DENSE DATA CENTER ARCHITECTURES

Recife

2020

KÁDNA MARIA ALVES CAMBOIM VALE

ANALYSIS OF DEPENDABILITY AND SUSTAINABILITY REQUIREMENTS TO SUPPORT THE DEPLOYMENT OF DENSE DATA CENTER ARCHITECTURES

A Ph.D. Thesis presented to the Graduate Program in Electrical Engineering of Federal University of Pernambuco in partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering.

Concentration area: Electronics.

Supervisor: Prof. Dra. Fernanda Maria Ribeiro de Alencar.

Co-supervisor: Prof. Dr. Jean Carlos Teixeira de Araujo.

Recife

2020

Catalogação na fonte
Bibliotecária Margareth Malta, CRB-4 / 1198

V149a Vale, Kádna Maria Alves Camboim.

Analysis of dependability and sustainability requirements to support the deployment of dense data center architectures / Kádna Maria Alves Camboim Vale. - 2020.

180 folhas, il., gráfs., tabs.

Orientadora: Profa. Dra. Fernanda Maria Ribeiro de Alencar.

Coorientador: Prof. Dr. Jean Carlos Teixeira de Araujo.

Tese (Doutorado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Elétrica, 2020.

Inclui Referências e Apêndices.

Texto em Inglês.

1. Engenharia Elétrica. 2. Data center. 3. Computação em nuvem. 4. Avaliação de dependabilidade. 5. Eficiência energética. 6. Sustentabilidade. 7. Análise de sensibilidade. I. Alencar, Fernanda Maria Ribeiro de (Orientadora). II. Araujo, Jean Carlos Teixeira de (Coorientador). III. Título

UFPE

621.3 CDD (22. ed.) BCTG/2021-35

KÁDNA MARIA ALVES CAMBOIM VALE

“ANALYSIS OF DEPENDABILITY AND SUSTAINABILITY REQUIREMENTS TO SUPPORT THE DEPLOYMENT OF DENSE DATA CENTER ARCHITECTURES”

A Ph.D. Thesis presented to the Graduate Program in Electrical Engineering of Federal University of Pernambuco in partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering.

Approved in: November 27, 2020

EXAMINATION BOARD

Prof. Dr. Fernanda Maria Ribeiro de Alencar
(Supervisor and Internal Examiner)
Universidade Federal de Pernambuco

Prof. Dr. Marco Aurélio Benedetti Rodrigues (Internal Examiner)
Universidade Federal de Pernambuco

Prof. Dr. Ricardo José Paiva de Britto Salgueiro (External Examiner)
Universidade Federal de Sergipe

Prof. Dr. Guilherme Nunes Melo (External Examiner)
Universidade Federal de Pernambuco

Prof. Dr. Érica Teixeira Gomes de Sousa (External Examiner)
Universidade Federal Rural de Pernambuco

To God!

ACKNOWLEDGEMENTS

I begin by thanking with the challenge of writing in a little space the name of everyone who is part of this achievement, who over the past six years have been with me or contributed in some way to me completing this stage.

To my friends of all hours, who I know I can count on your words and support in all situations, mainly, I thank the cats of BCC, Thaís, Maria, and Priscilla for the countless hours of shared experiences, tips on work, and maternity. I thank Jessica and Douglas for the moments of relaxation. Samara, Kedma, Tatiane, and Gabriel for the always powerful words for the most challenging moments. I also remember Séfora, Romário, Gabriela, Ricelly, and "Baby Luh" very fondly. I am grateful to Verusca, who battled with me with the same intensity, living and sharing this same challenge of do Ph.D. in times of pandemic; God and we know how much we had to face to overcome our limits.

On behalf of my coordinator, Rodrigo Gusmão, I would like to extend my thanks to my co-workers and my beloved institution UFRPE/UAG - currently UFAPE.

To the research group GRENCI (Group on Requirements Engineering, Network, and Computational Intelligence) that I had the opportunity to participate in the last few years, contributing and learning with colleagues, especially Aêda, Amora, Camila, Cinthya, and Timóteo.

To the research group MoDCS (Modeling of Distributed and Concurrent Systems), in the person of Paulo Maciel, I am grateful for the opportunity to start my first researches in this group. I wish to thank everyone who has endeavored to develop and improve the Mercury tool, which has skewed this work's results. I thank everyone who helped me many times, especially João, Carlos, Rubens, Jamilson, Rafael, Julian, Gustavo, Bruno, Danilo, and Erico.

To the Center of Technology and Geosciences (CTG) of the Federal University of Pernambuco (UFPE), to research provided support. To the current PPGGE coordinator, Juliano Bandeira, for his solicitude. The secretary Andrea and Cristina for their promptness.

I thank the professors of the examining committee for this work, Marco Benedetti, Ricardo Salgueiro, Guilherme Melo, Erica Sousa (Daniel Chaves, and Fernando Aires - as alternate members) for their availability to participate of this long-awaited moment. I am sure of how valuable their contributions were for improving this research, consequently, for my professional/personal growth.

I thank my advisor, Fernanda Alencar, for her contribution and patience over the years, for the valuable contributions given throughout the process. To my co-advisor, Jean Araujo, for passing on his teachings to me. I am grateful for having believed in my effort and the relaxed moments amid the ongoing search for good results.

To my group of renewal and union in Christ, which serves as an example of faith and hope. Thank you for your prayers and support. I thank Pedro and Amanda for their care in revising my text.

I am incredibly grateful to my beloved parents, Davi and Fátima, for the advice, education, and love offered. They understood my absence in so many family moments and who overcame the distance and the longing so that I could walk my paths. To my brothers, Kadmu and Kennedy, and my sisters-in-law, who especially this year presented me with my beloved nephews Laura and Lucas, thank you for all your support ever. To my uncles and aunts, cousins, mother-in-law, and father-in-law for all the advice and help.

With love, I thank my husband for the countless words of support for sustaining my moments of crisis and stress amid so many activities. To my baby Mateus, who with all his innocence taught me how strong I could be, especially for him, who undoubtedly suffered the most without understanding my absence in the countless calls of his "mom". I am grateful to the people who took care of him and my home because, without this assistance, I certainly would not have succeeded.

To God, I offer my eternal gratitude for the paths that allowed me to choose, for the strength, health, and courage to face each difficulty while doing this work. I am grateful for allowing all these people to be part of my paths. I am thankful for having managed to deal with the losses (in memory of my grandmother Luiza, my uncle, my aunts, and my cousins). Thank you for allowing me to achieve yet another victory.

The Lord is my strength and my shield; in him my heart trusts!
(Psalms 28:7)

ABSTRACT

The convergence of communication networks and the demand for storing and processing capacities of large amounts of information, especially in recent years, has driven requests for everything-as-a-service (XaaS) and has been generating, on an increasing scale, demands for new data centers (DC) infrastructures. Energy consumption per rack unit has increased due to the consolidation of spaces via server virtualization and as an effect of the density achieved by modern IT equipment. As these factors contribute to significant improvements, such as gains in availability and performance, they also contribute to global warming due to carbon dioxide (CO_2) emission, one of the leading greenhouse gases. This occurs both in the production of energy, which feeds these large infrastructures, and in the consumption of the entire IT load. Given this context, we analyze dependability and sustainability requirements to support data centers' implementation in this work. We initially performed a systematic literature review to identify possible problems in the area. We propose a set of models for power architectures and data center networks to quantify specific dependability and sustainability metrics. The proposed approach is based on the hybrid and hierarchical modeling for evaluating costs, exergy, electrical demand, efficiency, availability, reliability, etc., of the generated models. We propose energy flow models (EFM) to determine the minimum energy required for a data center project, considering the computer room (DC network). We created some scenarios to calculate the impact of environmental sustainability from raw material used in energy production. We also propose reliability block diagram (RBD) models to represent the power and network architectures of the data centers to identify the reliability importance of the components for the systems' availability. Besides, we use parametric sensitivity analysis techniques to carry out experiments in different scenarios to identify possible bottlenecks for architectural failure. Based on the formalism of the stochastic Petri nets (SPN), we propose a maintenance policy to quantify the impact caused on the availability of the networks in the face of the variation in the mean time to failure (MTTF) and repair (MTTR) of the components. Additionally, we check if the availability achieved reaches service levels specified for data centers given their tier classifications. The set of modeling techniques used serves as a subsidy for the experiments to represent real scenarios. Thus, we can quantify essential design requirements, identifying possible vulnerabilities to the system's failure before the implementation of a real architecture has been done.

Keywords: Data center. Cloud computing. Dependability evaluation. Energy efficiency. Sustainability. Sensitivity analysis.

RESUMO

A convergência das redes de comunicação e a demanda por capacidades de armazenamento e processamento de grandes quantidades de informação, principalmente nos últimos anos, tem impulsionado as requisições de "*everything-as-as-service*" (XaaS) e vem gerando, em escala crescente, demandas por novas infraestruturas de data centers (DC). O consumo energético por unidade de *rack* têm aumentado diante da consolidação de espaços via virtualização de servidores e como efeito da densidade alcançada pelos modernos equipamentos de TI. Na proporção em que esses fatores contribuem para melhorias significativas, como ganho em disponibilidade e desempenho, por exemplo, contribuem também para o aquecimento global, devido à produção de dióxido de carbono (CO_2), um dos principais gases do efeito estufa. Isso ocorre tanto na produção de energia, que alimenta essas grandes infraestruturas, quanto no consumo de toda a carga de TI. Diante desse contexto, nesse trabalho é feita a análise de requisitos de dependabilidade e sustentabilidade para apoiar a implantação de data centers. Inicialmente foi realizada uma revisão sistemática da literatura para identificação dos possíveis problemas na área. Na sequência, foi proposto um conjunto de modelos para arquiteturas de potência e de redes de data center para quantificar métricas específicas de dependabilidade e sustentabilidade. A abordagem proposta baseia-se na modelagem híbrida e hierárquica para a avaliação de custos, exergia, demanda elétrica, eficiência, disponibilidade, confiabilidade, etc., dos modelos gerados. Foram propostos os modelos de fluxo de energia (EFM) para determinar a energia mínima necessária para um projeto de data center, considerando a demanda de energia para a *computer room* (rede do DC). Foram criados alguns cenários para calcular o impacto da sustentabilidade ambiental diante do tipo de matéria-prima utilizada na produção de energia. Há ainda a proposta de modelos de diagramas de blocos de confiabilidade (RBD) para representar as arquiteturas de energia e de redes dos data centers a fim de identificar a importância da confiabilidade dos componentes para a disponibilidade dos sistemas. Adicionalmente, foram usadas técnicas de análise de sensibilidade paramétrica para realização de experimentos em diferentes cenários com a intenção de identificar possíveis gargalos para a falha das arquiteturas. A partir do formalismo das redes de Petri estocásticas (SPN), foi proposta uma política de manutenção para quantificar o impacto causado na disponibilidade das redes diante da variação dos tempos médios para a falha (MTTF) e reparo (MTTR) dos componentes. Complementarmente, foi verificado se a disponibilidade alcançada consegue atingir os níveis de serviços, que são especificados para os data centers dada suas classificações por camada. O conjunto de técnicas de modelagem utilizado serve de aporte para os experimentos realizados para a representação de cenários reais. Assim, torna-se possível quantificar requisitos importantes de projeto, identificando possíveis pontos de vulnerabilidades para a falha do sistema antes que a implantação de uma arquitetura real seja efetivada.

Palavras-chave: Data center. Computação em nuvem. Avaliação de dependabilidade. Eficiência energética. Sustentabilidade. Análise de sensibilidade.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

A	Availability
ATS	Automatic Transfer Switch
CO_2	Carbon Dioxide
CTMC	Continuous-Time Markov Chains
DC	Data Center
DoE	Design of Experiments
eDC	Enterprise Data Center
EFM	Energy Flow Model
g/kWh	grams per kiloWatt hour
IaaS	Infrastructure as a Service
iDC	Internet Data Center
IDC	International Data Corporation
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
IT	Information Technology
ITIL	Information Technology Infrastructure Library
kg/kWh	kilograms per kiloWatt hour
KooN	K-out-of-N (Block type RBD)
kW	kiloWatt
MTA	Mean Time to Active
MTBF	Mean Time Between Failures
MTTF	Mean Time to Failure
MTTR	Mean Time to Repair
NIST	National Institute of Standards and Technology

Opex	Operational expenditure
PaaS	Platform as a Service
PDU	Power Distribution Unit
PLDA	Power Load Distributed Algorithm
ppm	parts per million
PN	Petri Net
PUE	Power Usage Effectiveness
QoS	Quality of Service
RI	Reliability Importance
RBD	Reliability Block Diagram
ROI	Return on Investment
SaaS	Software as a Service
SLA	Service Level Agreements
SLR	Systematic Literature Review
SPN	Stochastic Petri Net
StArt	State of the Art through Systematic Review
STS	Static Transfer Switch
TCO	Total Cost of Ownership
UPS	Uninterruptible Power Supply
XaaS	Everything as a Service

CONTENTS

1 INTRODUCTION

1.1 CONTEXT AND MOTIVATION

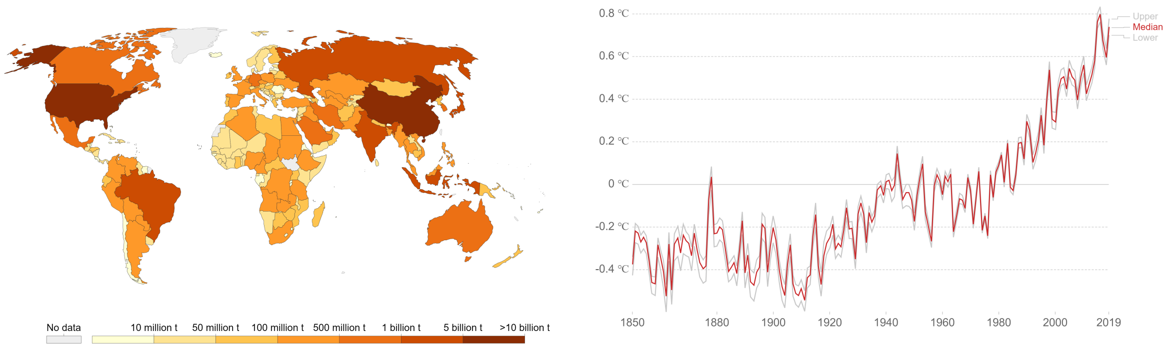
One of the biggest dilemmas in recent years has been related to the global emissions of greenhouse gases into the atmosphere due to burning fossil fuels (such as carbon, gas, and oil) that are mainly used for energy production. Global warming presents itself as one of the most urgent challenges today and is closely related to industrial activities and human actions in the environment. Figure 1 shows the total global greenhouse gas (GHG) emission versus the global average temperature anomaly.

Figure 1a shows the global emissions GHG, measured in "carbon dioxide-equivalents" (CO_2e) (RITCHIE; ROSER, 2020c). Countries like China and the United States (darker red) emitted more than 5 billion tonnes of GHG in 2016. Figure 1b shows how the planet warmed up. The red line represents the average annual temperature trend through time, with upper and lower confidence intervals shown in light gray (RITCHIE; ROSER, 2020a). We can see that the temperature in the year 1850 had a negative average (colder). This scenario has been changing to warmer temperatures, especially after the year 2000, where only a positive average can be seen.

Figure 1 – Global GHG emission X Global Temperature Anomaly.

(a) GHG Emission in 2016

(b) Global Temperature Anomaly



Source – (RITCHIE; ROSER, 2020c; RITCHIE; ROSER, 2020a)

GHG emissions increased significantly after the industrial revolution and have been increasingly driven by economic and population growth. Its effects have had direct impacts on global warming since the middle of the 20th century. As insignificant as it may seem, the increase in temperature by 1°C can bring some consequences through climate change; we can mention more accentuated ecological disasters such as floods, droughts, storms, heatwaves, physical and health problems, and interruption of water supply (RITCHIE; ROSER, 2020a).

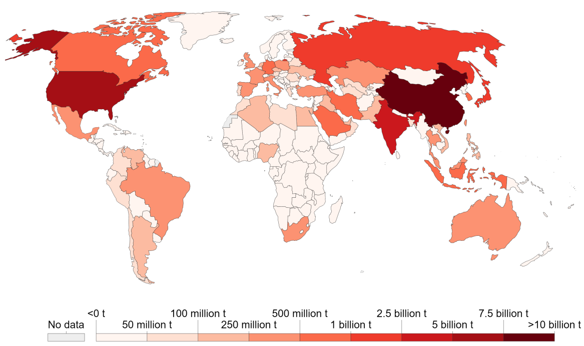
Carbon Dioxide (CO_2) is the main greenhouse gas. This is due to its ability to retain heat, calculated as "radiative forcing" (RF), in ice cores, the atmosphere, and other climate drivers. Methane, for example, has a more robust heat retention capacity than CO_2. However, as it is a much less abundant gas in the atmosphere, it makes CO_2 the most influential gas global warming (SCIENTISTS, 2020).

The industrial and academic community has made countless efforts to prevent the CO_2 concentration from reaching 450 ppm (parts per million), a critical limit for heating 2°C. In 2019, the world surpassed the 410 ppm mark (QUASCHNING, 2020a), and that value approached 415 ppm in CO_2 emissions, according to a survey presented by the Global Monitoring Laboratory (LABORATORY, 2020).

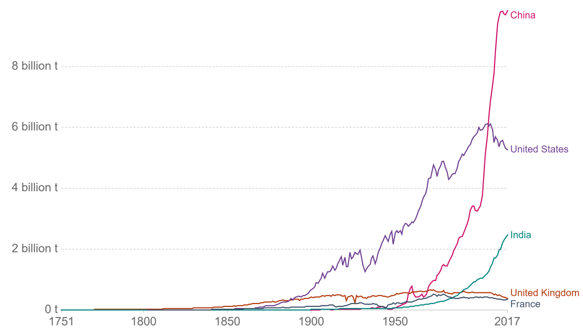
Figure 2 shows CO_2 emissions from the burning of fossil fuels for energy and cement production. Land-use change is not included. In Figure 2a, we can see CO_2 emissions for all countries in the world. Figure 2b shows emissions from China, the United States, India, the United Kingdom, and France through the timeline. It is notable the increasing CO_2 emissions from the 1990s, practically when industrialization started and had significant changes in living standards. The total sum of CO_2 emissions produced from fossil fuels and cement since 1751 is measured in tonnes.

Figure 2 – CO_2 Emissions by Country.

(a) CO_2 Emission in 2018



(b) CO_2 Emission by Country

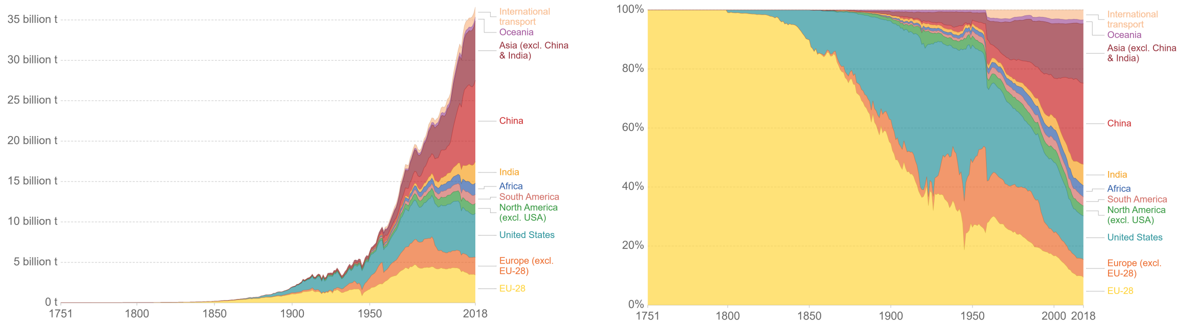


Source – (RITCHIE; ROSER, 2020b)

Figure 3 shows the change in the share of cumulative global CO_2 emissions by region over time - from 1751 to 2017. Figure 3a shows the cumulative CO_2 emissions over the years and each region's share. Note that Oceania and Asia emitted around 35 billion tonnes. In Figure 3b we see the percentage of CO_2 emissions by region. Note that until the 1900s, the United States (USA), which was the country that dominated the industrial market, is predominantly the region that most emitted. After that date, we see a strong contribution from other regions. This measures CO_2 emissions from fossil fuels and cement production only. Land-use change is not included (RITCHIE; ROSER, 2020a; RITCHIE; ROSER, 2020b).

Figure 3 – CO_2 Emissions by Region.

(a) CO_2 Emission by Region (tons) (b) CO_2 Emission by Region (%)



Source – (RITCHIE; ROSER, 2020a; RITCHIE; ROSER, 2020b)

We noticed an increase in CO_2 concentrations from the graphical presentations, which was no more than 300 ppm before the Industrial Revolution. It is a fact that there is a great inequality between emissions by countries or regions, which are proportionally related to economic activities and population quantity. Historically, the increase in CO_2 emissions, although having negative consequences, has been a by-product of positive improvements in humanity's conditions and lifestyle.

We can cite as an example the "cloud computing" that has reshaped the way companies use computing. The delivery of "everything-as-a-service" (XaaS), based on that technology, allows the usage of software, platforms, infrastructures, and databases by remote connection. Cloud computing is continuously expanding in the computing industry. It enables users to migrate their data to a remote location with minimal impact on system performance (WANG et al., 2010), promoting the use of computing infrastructure with different levels of abstraction and offering on-demand services, which are made available over the Internet (JANSEN; GRANCE, 2011).

The main drivers are business scalability, cost flexibility, and access to the industry for this technology. However, numerous challenges need to be overcome to ensure business continuity following the requirements established in the service level agreements (SLAs), among which are application compatibility, reliability, availability, governance, and compliance policies (NKHOMA; DANG; SOUZA-DAW, 2013).

According to a survey conducted by *International Data Corporation* (IDC) involving 263 IT executives to evaluate opinions on cloud services, the main challenges faced are related to security, with 87.5% of opinions, availability with 83.3%, and performance with 82.9% (POPOVIĆ; HOCENSKI, 2010). According to Yam et al. (2011), security and business continuity represent the most significant barriers, but other factors that stand out are related to relative advantage, uncertainties, geo-restrictions, the scope of the market, and support for external information technology.

In this context, organizations need to adopt multiple hypothetical relationships, such as risk-oriented cultures and normative and coercive pressures, to increase the way timely decisions are made (ALSHAMAILA; PAPAGIANNIDIS; LI, 2013). From a technical perspective, it is necessary to invest in improvements to increase security (e.g., reliability and confidentiality of data), availability, performance and scalability; within the organizational perspective, improvements must be made in terms of control, cost, standard, transparency and licensing.

Another survey conducted by IDC in 2016 claims that data centers (DC) and their digital business ecosystems (connectivity and cloud, which aggregate within them) are emerging as the main components of the supply chain and the creation of value in the digital economy (NAKKARIKE NAVEEN ANDA POSEY, 2016). In this sense, cloud storage is defined as a service based on opex (operational expenditure - capital used to maintain or improve a company's physical assets, such as equipment, properties, and real estate) deployed in public, private, or virtual data center, hosted outside a customer's physical DC to provide replication, archiving or backup services, applications or data (CHUTE, 2016).

Given this reality, there is an increase in the demand for computational energy, mainly driven by the large-scale construction of new DC infrastructures, which is responsible for storing/providing the mass of data produced and shared worldwide.

Maintaining high availability levels of large data centers has been an increasingly expensive task. Not only due to the general consumption of energy, which represents a total of 1% of all electricity consumed in the world (SVERDLIK, 2020), but mainly due to a large amount of CO_2 that is emitted when the choice of raw material used for energy production is a non-clean alternative.

Faced with the demand to maintain data centers' high availability, companies are being pressured to seek sustainable alternatives. Given this scenario, in which high availability rates must be achieved/provided by DC, at all times and around the world, it is imperative that the industry truly applies the concepts of sustainability (MURUGESAN, 2008) and best practices for the conscious use of energy, ranging from production to consumption.

The rising economic costs and environmental damage caused by expensive DC infrastructures have increased interest in sustainable computing, leading companies to a reality that forces them to be concerned with various aspects of sustainability. This concern can be caused by contractual requirements, law, or even by pressure from customers (VALE; ALENCAR, 2020).

Considering the financial investments made, the planning of the energy, cooling, and IT infrastructures is one of the significant challenges of the private cloud or DC (AHSON; ILYAS, 2010). Planning sustainable DCs can cost much more than conventional models making the return on investment (ROI) slow, a fact that leads companies to avoid these types of projects. Therefore, some initiatives focused on propagating techniques to make data centers more sustainable, considering IT equipment efficiency, energy efficiency, cooling efficiency, and clean energy alternatives. Other issues that hinder the planning process for implementing sustainable IT infrastructures are the lack of standardization, the lack of legislation in general, and the needs of sustainability, control, and environmental management.

Nowadays, energy projects for data centers need to consider the high consumption loads per rack due to the density achieved with modern IT equipment (MARIN, 2011). Capacity management aims to ensure availability when supplying electricity at a safe level. The available capacity is greater than the power used, without prejudice to future sizing, without burden to the total cost of ownership, but efficient enough to meet current needs.

The design, implementation, and maintenance of electrical, cooling, and data center IT architectures must also meet several dependability requirements to ensure the quality of service at a high level of reliability. One of the main challenges in a DC project is to achieve high availability, according to each layer's redundancy, mainly because the two main factors governing downtime are equipment design and maintenance philosophy. Achieving good repair times involves simplifying diagnosis and repair. Thus, designers must understand where the main bottlenecks are and the system's failure modes to work on maintenance strategies (SMITH, 2017). Low-reliability equipment must be more accessible and more easily removable to avoid further disturbance in maintenance or failure times.

Given the scope presented, this work proposes models for electrical architectures and IT architectures (networks) for data centers, according to redundancy specifications by tier. Our primary focus is on promoting the "green data center" concept, ranging from choosing clean energy sources to more efficient IT equipment. As a goal, we try to emphasize, especially for organizations and designers, that it is possible to achieve the levels of availability and reliability required for networks and, at the same time, meet ecological sustainability.

1.2 PROBLEMATIC AND OBJECTIVES

To enrich and strengthen the theoretical knowledge concerning the problem studied domain, we seek to answer a central research question based on a research hypothesis. From that point on, a systematic literature review (detailed in Chapter 4) was carried out in order to answer five research questions, which supported and guided the proposed solution. So, we have:

Research Question: How is it possible to identify bottlenecks in data center architectures to assist in the design phase, aiming to provide greater sustainability when dependability requirements are also met?

Hypothesis: The hierarchical and hybrid modeling approach makes it possible to find architectural bottlenecks to suggest improvements before implementing data centers.

The five questions that guided the SLR were the following:

- **Research Question 1 (RQ1) –** Is there an approach, technique, method, methodology, model, or a guide to support cloud computing migration?
- **Research Question 2 (RQ2) –** Is there an approach, technique, method, methodology, model, or a guide to support the migration to cloud computing, which considers non-functional requirements and sustainability?
- **Research Question 3 (RQ3) –** Which open problems related to the research area?
- **Research Question 4 (RQ4) –** Which are essential dependability requirements (attributes) for sustainable data center projects? (What are the biggest concerns?)
- **Research Question 5 (RQ5) –** Which models/modeling techniques are mostly used for proposing sustainable data center infrastructures?

Initially, only three research questions were prepared. Still, during the calibration phase, it was noticed that some search bases did not present results when looking for non-functional requirements and sustainability together, which resulted in the expansion of the research question RQ2 in two (See Chapter 4). After that, a new SLR was performed to include the questions RQ4 and RQ5, whose works found guided the proposed solution in the face of unresolved gaps. Besides, these works are presented as correlated to this research.

SLR reached 3,295 works, of which 29 compete for the first three questions, and 28 compete for the last two questions. All of them presented a more significant relation to the researched context. Of these 29 primary studies, 21 of them answered research question 3 (RQ3). In this way, it was possible to obtain a comprehensive and updated view of state of the art on the subject, categorizing the problems presented and identifying the gaps in the research context, which are briefly presented next (and more detailed in Chapter 4):

- Regarding migration to the cloud, among the most cited problems are problems with legislation and lack of standardization, in addition to difficulties in meeting quality or financial requirements, considering security aspects (such as virtualization vulnerability, data security, privacy, confidentiality), performance, availability, reliability, and costs.
- There is a lack of practical application concerning sustainability, culture still plastered, high cost of more efficient equipment, and delayed return on investment.
- About the challenges faced for the data center construction project, the complexity of the infrastructures was listed, composed of three major subsystems: energy, cooling, and networks; high energy consumption and inefficient cooling; the lack of physical space and the risks related to security, both physical and logical.

1.2.1 Objectives

Therefore, motivated by the shortcomings identified from the SLR, this research's main objective is to propose a hybrid and hierarchical modeling approach for the physical representation of electrical and networks (critical IT load) architectures for tiers 1, 2, 3, and 4 data centers. Thus, we can identify the sources contributing most to the emissions of CO_2 and these architectures' biggest bottlenecks. The methods offered aim to investigate dependability and sustainability requirements and verify the behavior of systems' availability against the different SLA classes, which were created to propose a maintenance policy for the data center networks.

Among the specific objectives of the research, we can list:

- To propose an energy demand to compute a data center load more efficiently based on the consumption by chassis racks[1].
- To build energy flow models to estimate dependability and sustainability metrics for electrical architectures.
- To build dependability models for assessing IT architectures.
- To investigate the environmental impact caused by different types of raw materials used for energy production that supplies the data centers represented.
- To assess the sensitivity of both electrical and IT architectures' components to identify those that have the most significant impact on availability.
- To assess the impact of a maintenance policy, which considers preventive and corrective maintenance approaches, on IT architectures' availability.
- To propose suggestions for improvement based on identifying the most critical components for each architecture.

1.3 JUSTIFICATION

According to Schulz (2016), IT services consumers are increasingly looking for solutions and products delivered via green supply chains and green ecosystems. Green supply chains extend beyond product logos with green backgrounds or pictures of green leaves on the packaging to make them feel good about going green. Regardless of stance or perception on green issues, the reality is that ecological and the corresponding economic aspects cannot be ignored for business and IT sustainability. Business benefits align with the most energy-efficient and low-power IT solutions combined with best practices to meet different data and application requirements in an economical and ecologically friendly manner.

[1] Rackmount chassis can accommodate 10 to 16 servers, depending on the manufacturer, thus saving space and costs.

Besides, given the high energy consumption of these infrastructures and the high amount of CO_2 emissions into the atmosphere, companies and designers must rethink the alternatives to get around these problems. In this sense, this work proposes an integrated approach of hierarchical and heterogeneous models to mitigate gaps in the lack of standardization for sustainable project solutions to implement green data center infrastructure.

> In data centers, the electrical and cooling systems' reliability and maintainability are fundamental design requirements to enable the successful operation of IT and cooling systems. It is possible to achieve the reliability goals and optimize energy efficiency simultaneously, but it requires close collaboration amongst the IT and facility teams to make it happen. Geng (2014, p. 32).

Nevertheless, we also offer experiments to present dependability metrics resulting from model evaluations and present the designed architectures' most sensitive components. We are also given more significant difficulties in achieving high availability.

1.4 NEGATIVE SCOPE

It is important to present what we cannot consider as the objective of this work. Thus, we determine the topics that are outside the scope of this research.

Firstly, it is not our goal to offer a solution that can be used to represent the plethora of possible physical configurations for electrical and IT architectures. We consider a simplistic scope to meet the minimum characteristics of redundancy, as suggested in the literature. However, the work can be adapted countless times to meet the aspects of several new architectures. It is essential to clarify that specific requirements for one organization are more prioritized than others and may not have been covered in this work. It is not up to us to implement all the possible attributes that meet organizations' different needs.

Second, although security has been presented as one of the main problems for adopting cloud computing, this topic does not fall within our scope.

Finally, it is not our goal to provide solutions for electrical and network connections' logical configurations. Our focus is on physical architectures.

1.5 ORGANIZATION

This thesis is structured as follows:

- In this Chapter 1, was presented the context and motivation to justify this work. The problem, the objectives, the proposed solution, and the negative scope were also shown.
- Chapter 2 presents the theoretical foundation, contextualizing the essential concepts for understanding the work.

- Chapter 3 displays the phases encompassed in the methodology of this research.
- Chapter 4 shows the systematic literature review. The entire process for its conduct is detailed and in the sequence is showing the results. At the end of the chapter are the works related to this research.
- Chapter 5 expose the models developed for the representation of electrical and IT (networks) infrastructures of four data center tiers, classified according to the redundancy specifications of the physical components.
- Chapter 6 presents the results of the evaluations and the experiments carried out.
- Chapter 7 shows the conclusions of this research, the limitations, the contributions, the publications, and the future works.
- After the references, are displayed the Appendices A, B and C.

2 BACKGROUND

This chapter presents the bibliographic basis to understand better the concepts and techniques addressed to carry out this work. Definitions are displayed for cloud computing and data center, dependability, sustainability, modeling techniques, maintenance policy, and service level agreements. At the end of the chapter, a set of technical standards aim to help companies and designers better plan their projects. The suggested topics are related to the research context studied.

2.1 CLOUD COMPUTING AND DATA CENTER

A widely accepted definition for cloud computing provided by the NIST (National Institute of Standards and Technology) refers to it as

> a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. (MELL; GRANCE et al., 2011, p. 2).

The characteristics achieved by this paradigm had long been foreseen in 1961 when *John McCarthy*, at a lecture by the MIT (Massachusetts Institute of Technology), said that computing would be like a public utility service (utility computing):

> "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility." (MAGOULÈS, 2009, p. 126).

These ideas were further explored by *Douglas Parkhill* in his book *The Challenge of the Computer Utility* in 1966 (PARKHILL, 1966).

Clouds have become large repositories of virtualized resources, such as hardware, software, and development platforms, easily accessible and dynamically configured to adjust to different workloads to optimize their use (VAQUERO et al., 2008).

The availability of these resources and their costs influenced the rapid growth of applications, increasing demand, and competition among providers to achieve a high quality of service. That resulted in the implementation of large data centers, which is expensive due to general energy consumption (KOOMEY, 2011).

The data center and its connections form the cloud's infrastructure, whether public or private (VERAS, 2012). In an internally hosted private cloud, the company needs to build its data center to work on the cloud. That includes hardware and software components, storage systems, power systems, management software, virtualization, physical access control, firefighting, precision cooling, power generators, and UPS (no-breaks) (MARIN, 2011).

According to Veras (2009) a data center comprises several interconnected assets, which provide capacity for large-scale data processing and storage for any organization type in a centralized manner. They are categorized into two groups: enterprise (eDC), built within the facilities of the company that uses it, and internet (iDC), which are responsible for cloud services for third parties (many still believe that data center only exists on network service providers).

Some project requirements must be considered for the implementation of these infrastructures: to meet the current and future demand of the business, to reduce the total cost of ownership (TCO), to facilitate the management of network, storage, and database resources (DBMS), consider support and maintenance, as well as security, performance, availability, downtime, and business flexibility. In this sense, large organizations, generally cloud providers, maintain more than one data center to distribute the data processing load and provide backup when needed, which can also serve as a security measure in the event of a disaster or failure in the primary data center (VERAS, 2009).

2.1.1 Redundancy Levels for Data Center Classification

The *Uptime Institute* classifies four different tiers for data centers, considering the redundancy in terms of architectural electrical, and mechanical (cooling) design. The more sophisticated the data center, the less downtime and the higher the cost of ownership (MARIN, 2011; VERAS, 2009). The tier classification was originated by *Uptime Institute* IV et al. (2006). It is currently used by several standards and associations and is found in many literature (MARIN, 2011; VERAS, 2009; HEISING et al., 2007; GENG, 2014).

There are four tier classifications, ranging from one to four, which classify a data center according to redundancy and fault tolerance characteristics. The satisfactory operation will depend on the integration of its subsystems (MARIN, 2011; VERAS, 2009).

- Tier 1 Data Center – It is the most basic classification model, in which there are no redundant cooling and electrical components or systems. A failure in electrical distribution can disrupt IT operations in whole or in part. Scheduled and unscheduled maintenance caused the site to stop. It has an availability of 99.67%, which is associated with a downtime of 28.8 hours per year (MARIN, 2011).

> Organizations or applications that use tier 1 facilities include small businesses or small office/home office (SOHO) locations that can tolerate some disruptions or downtime, which may rely on a managed service provider or hosting service for online presence and Web-based services. (SCHULZ, 2016, p. 130).

- Tier 2 Data Center – It has redundant components but only a distribution infrastructure for both energy and cooling. Failures in these systems can cause losses in all other parts.

Because it has a certain redundancy level, specific elements can perform maintenance activities without the site being completely shut down. It has an availability of 99.75%, and downtime can reach 22 hours per year (MARIN, 2011).

> Applications and data are covered by some form or level of business continuity and disaster recovery protection. Physical site selection is not as critical as for higher-tier data centers that provide additional availability, protection, and security. (SCHULZ, 2016, p. 131).

- Tier 3 Data Center – In this model, data centers must be self-sufficient, with the redundancy of components with the different routing of electrical circuits to meet the critical IT load. All essential IT equipment must be dual power supplies. Planned maintenance activities can be performed using redundant capabilities. Failures are more challenging to happen. Availability is 99.98%, and downtime can be 1.6 hours per year (MARIN, 2011).

> Tier 3 candidate applications or organizations include those with a worldwide presence, highly dependent on IT and associated servers, storage, and networking resources being available $7 \times 24 \times 365$. They have minimal to no tolerance for unscheduled service disruptions because of the high cost of downtime to business revenue, delivered service, image, or a combination of all of these. (SCHULZ, 2016, p. 131).

- Tier 4 Data Center – All wiring, devices, electrical, and climatic power must be redundant. The various electrical and climatic distribution paths can tolerate failures due to the different independent branches, all critical IT equipment having dual power supplies (at least). Availability is 99.99%, and downtime should not exceed 0.4 hours per year (MARIN, 2011).

> Organizations or applications that require tier 4 data center protection include large businesses, high-profile businesses, and businesses that are highly dependent on IT equipment and services being available. Applications that rely on tier 4 data centers can be independent of business revenue size or number of employees. The common theme for applications and businesses requiring tier 4 capabilities are the high cost of downtime in terms of lost revenue, lost productivity, spoilage or loss of raw goods, lost opportunity, public image and perception, regulatory fines, or support for essential services, including emergency dispatch, among others. (SCHULZ, 2016, p. 132)

According to Marin (2011), tier classifications are interpreted as levels of availability of the sites and reliability of the services offered. The application of tier classifications depends on on-site availability requirements, as well as their design guidelines.

Data center redundancy can be presented at different levels, based on the classifications of the *Uptime Institute* (IV et al., 2006; MARIN, 2011), they are:

- Basic data center (N) – No redundancy.
- Data center with $N + 1$ redundancy – has an additional unit concerning the primary data center.

- Data center with $N + 2$ redundancy – has two additional units concerning the minimum required.
- Data center with $2N$ redundancy – has two additional units, modules, distribution routes, or systems for each referred by a primary site.
- Data center with redundancy $2(N+1)$ – has two complete $N+1$ units, modules, distribution routes, or systems. Even in the event of failures or maintenance, site operations will not be interrupted.

The lowest ranking element will determine the overall tier classification. If a high rating is required, then the electrical design (and each associated subsystem) must meet that rating (MARIN, 2011).

> The electrical distribution system in a data center encompasses numerous equipment and subsystems that begin at the utility and building transformers, switchgear, UPS, PDUs, and power supplies, ultimately powering the IT's fans and internal components equipment. (GENG, 2014, p. 32).

There are specific design parameters that can assist in sizing and choosing the tier classification. In addition to the operational side, it is necessary to think about TCO and ROI strategies. Another point that cannot be overlooked in the project is related to financial losses given a certain downtime period. Following are some recommendations for the electrical design of data centers in accordance with the tier redundancy specifications from Marin (2011):

- Tier 1 – The most basic classification model, in which there is no redundancy in the physical and logical energy routes. It has one or a group of generators to take charge of the site due to a lack of power from the concessionaire. An ATS (Automatic Transfer Switch) is typically used to switch the distribution system to the generator in a utility failure event. It must have a UPS (Uninterruptible Power Supply) and PDU (Power Distribution Unit) to power the IT load.
- Tier 2 – UPS and cooling systems must provide redundant modules for $N + 1$ (at least). An electrical generator system is required to supply the entire load. Failures in the electrical and cooling systems can cause failures in all other components. It must meet all tier 1 requirements so that a failure or maintenance service on a feeder or electrical distribution board can cause the DC to stop. Have one or a group of generators, ATS, PDU, and additional PDUs are recommended.
- Tier 3 – Must meet the requirements of tier 2 and must be self-supporting, where $N + 1$ (at least) redundancy for the distribution paths, modules, UPS, generators, and electrical feeders. This redundancy can be met through two electrical power sources for each CRAC (Computer Room Air Conditioner) unit.

- Tier 4 – Must meet all tier 3 requirements. All wiring, devices, and electrical supply must be redundant. Electrical and mechanical systems must have $2(N+1)$ redundancy, with at least two alternative energy sources. Understand sources of energy from utilities, but the DC itself can maintain that. As an example, we have a solar plate circuit. It has a generator set with a minimum recommended configuration of $2(N+1)$, ATS, UPS, PDU, and additional PDUs are recommended on the secondary circuit.

> All of these components will have a degree of inefficiency, resulting in a conversion of the electricity into heat (energy loss). Some of these components have a linear response to the percent of total load they are designed to handle; others will demonstrate a very nonlinear behavior. (GENG, 2014, p. 32).

2.1.2 ANSI/TIA/EIA-942

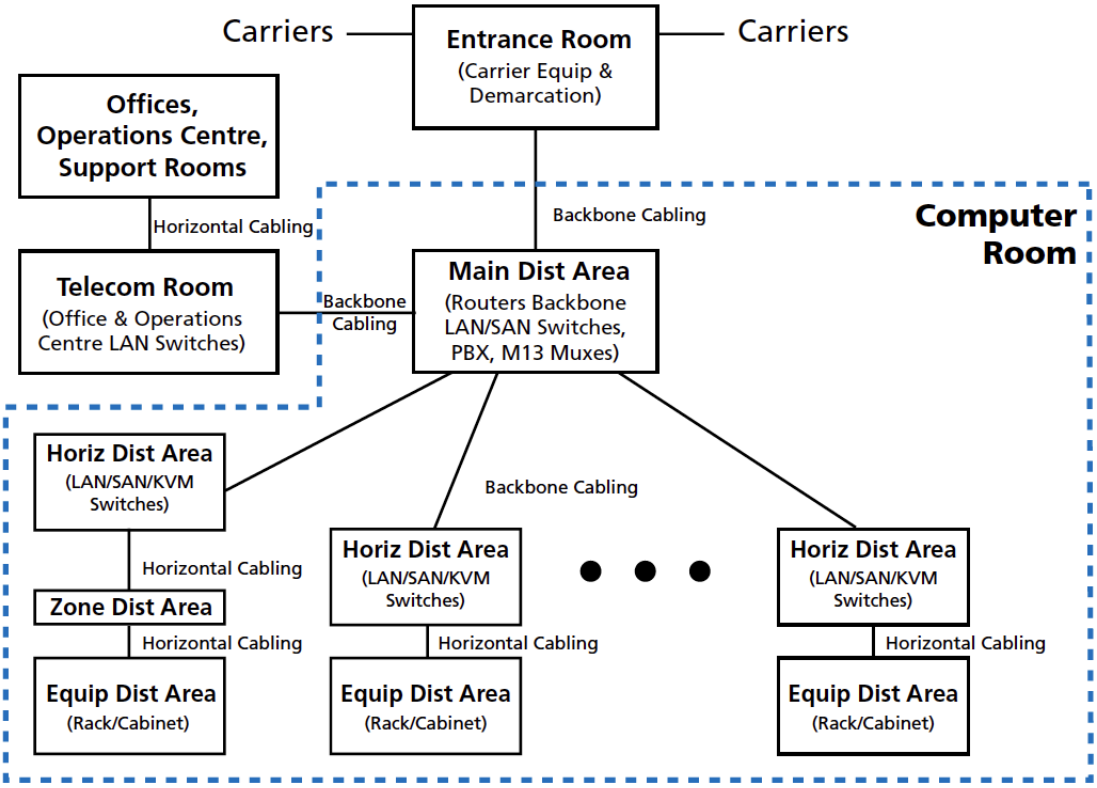
The ANSI/TIA/EIA-942 standard - Telecommunications Infrastructure Standard for Data Centers – is based on other standards that guide the design of the different subsystems, but it is the main one for any DC project. The specifications for the various critical systems include the design of the installations, the network, the structured cabling, the automation systems, power supply, firefighting, and refrigeration control (DIMINICO; JEW, 2006; VERAS, 2009).

The basic topology of a data center can be described in several ways. According to ANSI/TIA/EIA-942 standard, which guides design requirements from construction to activation, it can be categorized by layers or zones, making it easier to compare data centers, cost, and performance estimates (DIMINICO; JEW, 2006; VERAS, 2009). Figure 4 illustrates DC's topology, according to the ANSI/TIA/EIA-942.

The main components and their functionalities are described next, according to DiMinico e Jew (2006), Veras (2009):

- ER (Entrance Room): It is the interconnection space between the structured cabling of the data center and the cabling coming from the connectivity providers.
- MDA (Main Distribution Area): It is the main point of distribution of structured cabling in the data center where the primary connection equipment (core) routers, Ethernet switches, SAN switches, PABX are located.
- HDA (Horizontal Distribution Area): It is an area used to connect EDA equipment to MDA through intermediate equipment and horizontal cabling.
- EDA (Event-Driven Architecture): Area for installing racks, terminal equipment (servers, firewalls, storage, appliances), and data or voice communication equipment.
- ZDA (Zone Distribution Area): It is an optional point of interconnection of the horizontal cabling between the EDA and the HDA, allowing a quick and frequent configuration, usually positioned under the floor.

Figure 4 – Topology of a Data Center – ANSI/TIA/EIA-942.



Source – (DIMINICO; JEW, 2006)

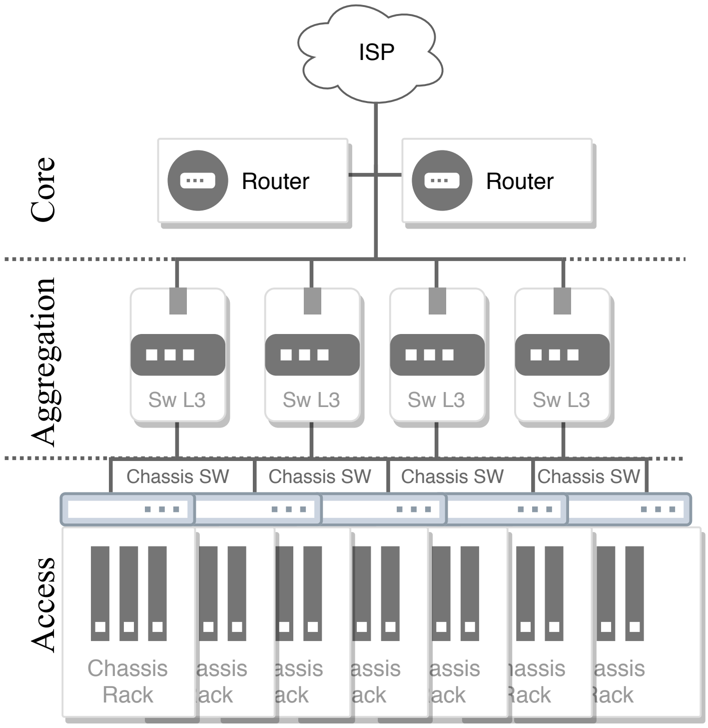
2.1.3 Layered Architecture

The architectural model based on layers is considered to facilitate a data center's design goals, such as performance, flexibility, scalability, resilience, and management (VERAS, 2009). Figure 5 presents the three-layer model that is commonly used in DC networks. This type of network modularization considers both the physical and the logical design. In the access layer are the chassis racks that represent the network storage endpoints. The aggregation layer links the access and core layers. The core layer represents the high-speed connection point between the Internet Service Provider (ISP) and external networks for incoming and outgoing data flow (RAZA; TURNER; ASAD, 2000). This image was created based on this work's network components.

2.1.4 Computational System

Currently, organizations are prioritizing more efficient projects to achieve savings in space, resources, and investment. These requirements are strongly desired, especially for data centers that aim to adopt the "green" seal - sustainable. In this sense, the choice of efficient components for all subsystems, which do more work per watt of energy (SCHULZ, 2016), should be privileged.

Figure 5 – Three-layers Network Model.



Source – Elaborated by the author.

A computing platform (hardware, firmware, and software) that runs an operating system (OS) collects applications to perform various operations, including file systems, logical volume manager, and device drivers. That is also called a computational system. Examples are desktops, laptops, servers, and mobile devices, consisting of processing, storage, input, and output devices. Computer systems used in building data centers and virtualized infrastructures are usually classified into four categories (IV et al., 2006):

- Tower computing system – it is built in a vertical enclosure called "tower", which is similar to a table cabinet.
- Rack-mounted computing system – designed to be attached to a frame of a so-called rack, or coulis, which is a cabinet with standardized measures containing several mounting slots called "bays".
- Blade center systems – known as a blade server, is an electronic circuit board containing only central processing components such as processor, memory, network controller, storage unit, and input and output ports. Blade servers are being widely adopted due to their modular design, which provides framework technology for multiple servers, facilitating power-sharing, cooling, and other shared services within the chassis. Integrated network switches provide additional space conservation and significant reductions in cabling (Smith et al., 2008). Compared to server racks, blades can be more easily managed due to

the density and component grouping characteristics (LEIGH; RANGANATHAN, 2007). Figure 6 presents the blade center system, which is used in this work to represent the access layer components.

Figure 6 – Blade Center Systems.

Source – (STAS, 2020)

- Large computing system – mainframes are the primary representative, which has high availability, scalability, security, and energy efficiency.

2.2 DEPENDABILITY REQUIREMENTS

The infrastructure of a data center should generally offer a minimum availability of 99.67% (tier 1), but to support highly reliable systems, the ideal would be that this number should be at least 99.9%, considering the capacity of withstanding failures. Each additional "nine" increases the order of magnitude of availability by a factor of 10. Thus, an increase from 99.99% to 99.999% is quite significant and represents a tenfold cost in investment (MARIN, 2011).

Given the accelerated growth of the digital economy, data center projects must promote sufficient resources to meet high availability. In this sense, a range of metrics can be quantified. Among the various metrics for the most different purposes is the dependability attributes that allow obtaining quantitative measures, often crucial for analyzing other systems types' services. Some features commonly used in systems reliability and dependability analysis are referenced in several publications (HEISING et al., 2007; TRIVEDI et al., 2009; KUO; ZUO, 2003; MACIEL; TRIVEDI; KIM, 2010). The following are those that are of interest to this research.

- Component – A piece of electrical or mechanical equipment viewed as an entity to reliability analysis (HEISING et al., 2007).
- Failure – The end of a component or system's ability to perform a necessary function (HEISING et al., 2007).

- System – A group of components connected or associated in a fixed configuration to perform a specified function (HEISING et al., 2007).
- Mean time to failure (MTTF) - The mean exposure time between consecutive repairs (or installations) of a component and the next failure of that component (KUO; ZUO, 2003). MTTF is commonly found for nonrepairable items, such as fuses or lamps. However, it is common to use this metric for reliability analysis because the modeling properties are static (HEISING et al., 2007). Normally, the manufacturer can present millions of hours until failure in the component datasheet, which can correspond to thousands of years. But,

> these data are usually determined based on accelerated tests (stress), which make certain assumptions about the operating conditions under which the components will be used (for example, that the temperature will always remain below limit), workloads, and duty cycles or on-time patterns, and that individual data center handling procedures are followed. In practice, operating conditions may not always be as ideal as assumed in the tests used to determine data sheet MTTFs. (SCHROEDER; GIBSON, 2007, p. 8)

MTTF of a system can be calculated by Equation 2.1.

$$MTTF = \int_0^\infty R(t)dt \tag{2.1}$$

- Mean time between failures (MTBF) – the average time between system failures. The hardware manufacturer generally supplies MTBFs. MTBF is obtained by Equation (2.2).

$$MTBF = MTTR + MTTF \tag{2.2}$$

> Technically, MTBF should be used only for a repairable item, while MTTF should be used for non-repairable items. However, MTBF and MTTF are commonly used for both repairable and nonrepairable items. (GENG, 2014, p. 277).

- Mean time to repair (MTTR) – The mean time to replace or repair a failed component (KUO; ZUO, 2003). The logistics time associated with the repair, such as purchasing parts, mobilizing the team, is not included (HEISING et al., 2007). MTTRs are closely related to the maintenance policy adopted by and can be achieved by Equation 2.3.

$$MTTR = MTTF \times \frac{UA}{A} \tag{2.3}$$

Where UA represents system downtime (Equation 2.4) and A represents system availability (Equation 2.7).

$$UA = 1 - A \tag{2.4}$$

- Mean down time (MDT) – the average time the system is non-operational, for any reason, e.g., corrective maintenance activities, preventive maintenance activities, and unavailability of resources(KUO; ZUO, 2003).

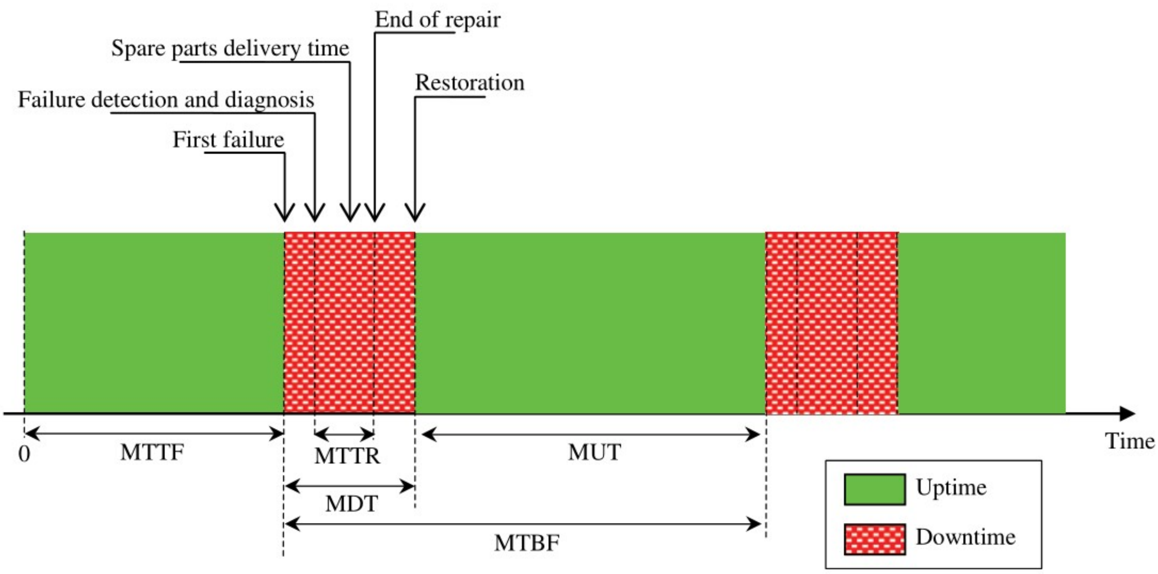
$$MDT = FDT + SDT + MTTR \tag{2.5}$$

  Where FDT is the fault detection time, and SDT is the spare delivery time.

- MUT – It is the mean uptime.

Figure 7 shows the equipment state as a function of time (GENG, 2014, p. 277). Time to failure and downtime are random variables. The mean values are defined as mentioned above.

Figure 7 – Equipment State During its Service Period.



Source – (GENG, 2014, p. 277)

- Availability – The ability of an item, under combined aspects of its reliability, maintainability, and maintenance support, performs its required function at a given time or for a given period. In Trivedi et al. (2009), the authors define availability as the system's ability to perform its intended function at a specific time. That is the probability that it will be operational for a certain period or that it will have been restored after the occurrence of a failure event.

  – Inherent availability (Ai): The instantaneous probability that a component or system will be up or down. Ai considers the only downtime for repair to failures. No logistics time and preventative maintenance are included (HEISING et al., 2007).

- Operational availability (Ao): The instant probability that a component or system will be active or inactive but differs from Ai in that it includes all downtime. All the delays for scheduling, travel time, parts, and others are included if it takes 24 h to fly a part in to repair the equipment, that adds to the repair time (HEISING et al., 2007).

The availability can be expressed in terms of the number of nines according to Equation 2.6 (MACIEL; LINS; CUNHA, 1996).

$$N = 2 - \log(100 - A) \tag{2.6}$$

Where 100 represents the maximum availability level that the system can achieve.

To obtain the availability of a system by steady-state analysis or simulation, the Equation 2.7 is used (MACIEL; LINS; CUNHA, 1996).

$$A = \frac{MTTF}{MTTF + MTTR} \tag{2.7}$$

- Reliability – The reliability of an item or system is the probability that a system satisfactorily (or adequately) fulfills its specific purpose, for a given period, until the occurrence of the first failure (MACIEL; TRIVEDI; KIM, 2010), (XIE; POH; DAI, 2004). I.e., it is the probability that the system will not fail until the specified time (KUO; ZUO, 2003), (RAUSAND; HøYLAND, 2004).

> Reliability engineering is defined as the science of failure. The first issues and reliability concepts appeared at the beginning of the twentieth century. Today, reliability engineering is widely used in many areas. Reliability engineering uses equipment reliability statistics, probability theories, system functional analysis, and dysfunctional analysis to set requirements, measure or predict reliability, identify system weakness points, and propose improvements. (GENG, 2014, p. 275).

Mathematically, the reliability function **R(t)** is the probability that the system will function correctly, without failure, in the time interval from 0 to t, as presented in Equation 2.8 (MACIEL; TRIVEDI; KIM, 2010).

$$R(t) = P(T > t), t \geq 0 \tag{2.8}$$

Where T is a random variable, representing the time to failure.

The failure probability, or inverse of reliability, is then represented by Equation 2.9, which is known as the T distribution function (MACIEL; TRIVEDI; KIM, 2010).

$$F(t) = 1 - R(t) = P(T \leq t) \tag{2.9}$$

- Maintainability – When a system goes down, repair operations are usually applied to correct the failure. The system, then, recovers its operational state due to adjustments made or even component replacements (XIE; POH; DAI, 2004). From this perspective, maintainability is the system's ability to undergo modifications and repairs after a failure event in a certain period (TRIVEDI et al., 2009). Maintainability is described by Equation 2.10, where T denotes repair time or total downtime. This equation represents maintainability since the repair time T has a density function $g(t)$ (TRIVEDI et al., 2009).

$$V(t) = P\{T \leq t\} = \int_0^t g(t)dt \tag{2.10}$$

2.3 SUSTAINABILITY

The concept of sustainability addresses human beings' impact on nature and the technological capacity to produce more durable materials that generate less waste, avoiding the planet's degradation process. This process has always occurred, but it has been accelerating due to globalization and increasing societies' consumption. The word sustainability derives from the term "sustainable societies" consumption recognized internationally in 1972, at the United Nations Conference on the Human Environment, held in Stockholm, Sweden. Sustainable development meets the needs of the present without compromising future generations' ability to meet their own needs (WCE&D, 1987).

Paul James, in his book *Urban Sustainability in Theory and Practice: Circles of Sustainability*, states that sustainable development interconnects four significant domains: ecology, economics, politics, and culture. He says that a truly sustainable system takes these domains into account, making the result ecologically positive, economically viable, politically correct, and culturally accepted by society (JAMES, 2014).

The concept widely publicized in recent times for sustainability and sustainable development is supported by the economic, social, and environmental pillars. In general, a process or a system can be said to be sustainable when it presents conditions for its permanence, at a certain level, for the longest possible period, paying attention to the conditions of economic, social, and environmental viability.

However, a significant cultural change is required, so the discussion focuses on how technology can enable efficient, durable, and less harmful goods to the environment. The production of this type of goods requires investments, mainly in research and technological innovation, changes in standardization, changes in the relationship with customers, in the production chain, in legislation, among other factors, which challenge companies to function transparently and responsibly (HART; MILSTEIN, 2004).

The best corporate practices that regulate environmental responsibility meet Green IT governance requirements and benefit stakeholders. Among the specific governance standards for companies, there is the Global Report Initiative (GRI) and the International Organization for Standardization (ISO) with norm 14.001, which guides, specifies, and helps in the improvement of an environmental management system (MAKOWER, 2009). Greenpeace also regulates and recognizes environmentally friendly companies.

The challenge is to find practical solutions to reduce greenhouse gas emissions (GHG), which absorb radiation in the infrared frequency, trapping heat in the atmosphere, thus contributing to global warming. Due to the increasing development of industrial processes, including ICTs and data centers, there was an increase in energy consumption (RICHTER, 2013), partly responsible for a portion of CO_2 emissions.

In this sense, the commitment to sustainability, the use of ecologically responsible practices, alternative and clean energies, and other issues that corroborate with sustainable development open numerous opportunities for research and innovation. However, companies and users must apply these concepts so it is possible to reduce costs, energy consumption, and environmental damage (MURUGESAN, 2008).

According to Schulz (2016), specific problems or combinations of problems vary according to the organization's size, location, reliability, and complexity of IT applications and servers, among other factors. In general, IT organizations have not prioritized being perceived as green, focusing on apparently non-ecological issues of PCFE (energy, cooling, physical space, and environmental health and safety).

Part of the green gap is that many IT organizations address (or need to resolve) PCFE problems without making the connection that they are adopting green practices, directly or indirectly. As a result, industry messages do not effectively communicate the availability of green solutions to help IT organizations solve their problems. When dealing with current IT problems, including energy, cooling, and physical space, along with the disposal and recycling of assets, the by-products are economical and ecologically positive. Likewise, the shift in thinking from avoiding energy to more efficient energy use helps both economically and ecologically (SCHULZ, 2016).

2.3.1 Energy Efficiency and Sustainability for Data center

Data centers require more power, cooling, and physical floor space to accommodate the servers, storage, and network components necessary to support growing demands. However, there are many challenges about power, cooling, floor space, and greenhouse gas emissions. For example, the energy efficiency gap is the difference between the amount of work accomplished or information stored in each footprint and the energy consumed. In other words, as more prominent the energy efficiency gap, as better (SCHULZ, 2016).

Next, we describe some metrics used to quantify sustainability in data centers regarding energy efficiency, costs, and CO_2 emissions.

- Exergy - It is a thermodynamic metric that represents the maximum amount of energy present in a system that is available to be converted into useful work (CALLOU et al., 2013a; DINCER; ROSEN, 2012). The efficiency of a system's energy consumption can be estimated from the exergy used. Consider F as a quality factor represented by the Exergy/Energy ratio so that exergy can be calculated with the Equation 2.11.

$$\text{Exergy} = Energy \times F \tag{2.11}$$

For example, the exergy present in a gallon of oil is greater than the exergy present in the same amount of water at room temperature. The oil can generate work (for example, moving cars); however, water at room temperature does not (CALLOU et al., 2013a).

- Power Usage Effectiveness (PUE) - It is one of the metrics established to assess the energy efficiency of data centers (SCHULZ, 2016). PUE is defined as the total data center load divided by the useful critical load, according to Equation 2.12.

$$\text{PUE} = \frac{\text{DC}_{load}}{\text{IT}_{load}} \tag{2.12}$$

Where DC_{load} is the total load of the data center infrastructure, in kW and IT_{load} is the effective critical load, in kW.

- Operational Cost - For this work, we consider the data center energy infrastructure's acquisition costs and operational cost. The operating cost considers the DC operation period, the energy consumed, energy cost, and the availability (according to the tier classification). Equation 2.13 denotes the operating cost (FERREIRA et al., 2018).

$$OP_{Cost} = (P_{input}) \times (Energy_{Cost}) \times (T) \times (A + \alpha(1 - A)) \tag{2.13}$$

Where P_{input} is the power supply input, $Energy_{Cost}$ is the energy cost per unit of energy, T is the period considered, A is the system availability, α is the percentage of energy that continues to be consumed when the system fails.

- CO_2 Emission - We also calculate the environmental sustainability impact by quantifying the amount of CO_2 produced by the data center's operation (although the Equation 2.14 can be used for other contexts). Different types of raw materials for energy generation will be considered. Its percentage of use is multiplied by its impact factor. For example, suppose that D_e represents the DC energy demand, $_i$ represents the energy source used, P_i represents the percentage of the energy source, and F_i represents the factor of aggression to the environment. Thus, the calculation for CO_2 emission, is given by Equation 2.14 (FERREIRA et al., 2018).

$$CO_2Emission = \sum_{i=1}^{n}(D_e \times P_i \times F_i) \tag{2.14}$$

2.4 MODELING TECHNIQUES

The representation of systems through modeling techniques allows obtaining useful information about their structure and behavior. The modeling enables the abstraction of the system's dynamic behavior levels of representation. Thus, in most cases, prototypes of models are generated, combined to achieve such representation. Some examples are the models of performance, reliability, availability, costs, among others. The modeling techniques can be classified as analytical techniques and techniques based on simulation.

Analytical modeling techniques use a set of equations and mathematical functions to describe the system's behavior, considering specific parameters that can be adapted to others, making it possible to verify the relationship of the application's effects by the parameters defined in the equations. The simulation techniques are based on abstract models, capable of representing the system's essential characteristics so that the values assumed by parameters can be more efficiently controlled (LILJA, 2000).

> The modeling process allows for considerable flexibility in the choice of state variables, as is the case in selecting input and output variables. In other words, there is no unique state representation for a given system. Usually, however, intuition, experience, and natural physical quantities that serve as state variables lead to good state-space models. (CASSANDRAS; LAFORTUNE, 2008, p. 9)

Modeling techniques may require defining attributes and metrics that serve as input to perform the calculation for possible analyzes. Among the various modeling methods are the performance and dependability models through techniques like Fault Trees, Markov Chains, Reliability Block Diagram (RBD), Achievability Graph, Petri Nets (whether Color (CPN) or Stochastic (SPN)), and Energy Flow Model (EFM). Among the various tools available for creating models stand out: Sharpe (SAHNER; TRIVEDI, 1987), Block-Sim (RELIASOFT, 2012), TimeNet (GERMAN et al., 1995a), GreatSPN (CHIOLA et al., 1995), SPNP (CIARDO; MUPPALA; TRIVEDI, 1989), and Mercury (MACIEL et al., 2017).

The following are the concepts and definitions of the techniques for generating the models in this study: EFM, RBD, and SPN.

2.4.1 Energy Flow Model (EFM)

It is proposed to estimate costs, exergy, the necessary electrical demand, and other sustainability metrics for data centers, respecting the restrictions of use and efficiency of the components. This is done with algorithms that run through EFM and calculate these metrics (MACIEL et al., 2017).

The EFM is a directed acyclic graph defined in (CALLOU et al., 2013a; FERREIRA et al., 2020) as follows:

$G = (N, A, w, f_d, f_c, f_p, f_\eta)$, where:

- $N = N_s \cup N_i \cup N_t$ represents the set of nodes (*i.e.*, the components), in which N_s is the set of source nodes, N_t is the set of target nodes, and N_i denotes the set of internal nodes, $N_s \cap N_i = N_s \cap N_t = N_i \cap N_t = \emptyset$;
- $A \subseteq (N_s \times N_i) \cup (N_i \times N_t) \cup (N_i \times N_i)$ = {(a,b) | a $\neq$ b} denotes the set of edges (*i.e.*, the component connections).
- $w : A \to \mathbf{R}^+$ is a function that assigns weights to the edges (the value assigned to the edge (j, k) is adopted for distributing the energy assigned to the node, j, to the node, k, according to the ratio, $w(j,k)/\sum_{i \in j^\bullet} w(j, i)$, where $j^\bullet$ is the set of output nodes of j);
- $f_d : N \to \begin{cases} \mathbf{R}^+ & \text{if } n \in N_s \cup N_t, \\ 0 & otherwise; \end{cases}$
  is a function that assigns to each node the heat to be extracted (considering cooling models) or the energy to be supplied (regarding power models);
- $f_c : N \to \begin{cases} 0 & \text{if } n \in N_s \cup N_t, \\ \mathbf{R}^+ & otherwise; \end{cases}$
  is a function that assigns each node with the respective maximum energy capacity;
- $f_p : N \to \begin{cases} 0 & \text{if } n \in N_s \cup N_t, \\ \mathbf{R}^+ & otherwise; \end{cases}$
  is a function that assigns each node (a node represents a component) with its retail price;
- $f_\eta : N \to \begin{cases} 1 & \text{if } n \in N_s \cup N_t, \\ 0 \leq k \leq 1, k \in \mathbf{R} & otherwise; \end{cases}$
  is a function that assigns each node with the energetic efficiency.

Some metrics capable of estimating sustainability impacts are presented after the evaluation of an EFM model. For example its lifetime exergy (available energy) consumption (MACIEL et al., 2017).

2.4.2 Reliability Block Diagrams

It is essential to establish what constitutes a failure because only then can it be determined which failure modes, at the component level, actually render the system inoperable. The reliability block diagram (RBD) modeling technique allows the representation of systems as a set of functional blocks in series, parallel, bridge, or K-out-of-N (KooN) interconnected according to each blocks' failure effect in the overall system reliability (SMITH, 2017). Input parameters include MTTF and MTTR.

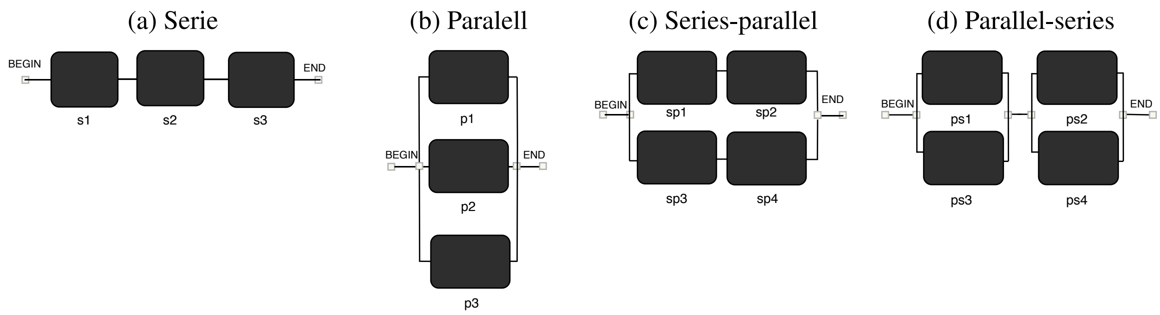
RBDs are adopted to represent block structures representing systems and subsystems with components connected according to their specified functions, or a trust relationship (KUO; ZUO, 2003; XIE; POH; DAI, 2004; RAUSAND; HøYLAND, 2004). RBDs are mainly used in modular systems that consist of many independent modules and contain an input (or source - on the left) and an output (or destination - on the right). To identify reliability by building RBD models, some tools as Mercury (MACIEL et al., 2017) and Blocksim7 (RELIASOFT, 2012) are used.

For series diagrams, all components must work. For parallel diagrams, only one component needs to work for the system to be operational (TRIVEDI; MALHOTRA, 1993). KooN blocks represent structures in which the subsystem can function if *k* or more components are in the operational state (XIE; POH; DAI, 2004). For example, a structure in which there are five components and three are needed to provide the expected service forms a *3-out-of-5* structure. Series and parallel structures are special cases of KooN, a series structure is an *N-out-of-N*, and a parallel structure is a *1-out-of-N* (KUO; ZUO, 2003; SAHNER; TRIVEDI, 1987).

For the mathematical definition of this logical arrangement's reliability, it is necessary to define the discrete random variable X, which defines the number of blocks that do not present failures in a given time interval. The probabilistic dependability events are independent for each block of the K configuration of N, and all N blocks have the same failure rate (XIE; POH; DAI, 2004).

Figure 8 presents some possible configurations for RBDs. Figure 8a presents a system with a connection of blocks in series. Figure 8b presents a parallel system. Blocks connected in series and blocks connected in parallel can be combined to form a new block. Figure 8c shows a series-parallel combination. Figure 8d shows a parallel-series combination.

Figure 8 – Reliability Block Diagrams



Source – Elaborated by the author.

Following are presented the equations for the different reliability calculations of the systems. The equations 2.15 to 2.19 were obtained in (TRIVEDI; MALHOTRA, 1993). The reliability of two blocks connected in series is obtained through Equation 2.15.

$$R_S = R_1 \times R_2 \tag{2.15}$$

Where R_1 describes the reliability of block 1, and R_2 represents the reliability of block 2.

Equation 2.16 calculates the reliability for series structures with n components. Suppose that $R_i(t)$ corresponds to the reliability of the block b_i at time t.

$$R_s(t) = \prod_{i=1}^{n} R_i(t) \tag{2.16}$$

The reliability of two blocks connected in parallel is obtained through Equation 2.17.

$$R_P = 1 - \prod_{i=1}^{2}(1 - R_i) \tag{2.17}$$

The reliability of n blocks connected in parallel is obtained through Equation (2.18).

$$R_P = 1 - \prod_{i=1}^{n}(1 - R_i(t)) \tag{2.18}$$

Where $R_i(t)$ corresponds to the reliability of block b_i at time t.

Equation 2.19 calculates the reliability of non-identical KooN blocks.

$$R_s = \sum_{r=i}^{n} \binom{n}{r} R^r (1 - R)^{n-r} \tag{2.19}$$

Where n is the total number of non-identical blocks, r is the minimum number of units required for system success. R is the reliability of each unit.

2.4.2.1 Sensitivity Analysis

Sensitivity analysis is a technique used to determine the most relevant factors to the measurements or outputs. In the systems dependability analysis, it is possible to apply sensitivity analysis to identify the components that most influence the availability, reliability, or performance (HAMBY, 1994).

It can be used in analytical models, from the calculation of partial derivatives of interest metrics, to provide a unique sensitivity coefficient to determine the most influential factors in the model results (FRANK; ESLAMI, 1980; HAMBY, 1994). This approach can find performance or availability bottlenecks in the system, thus driving improvement and optimization (BLAKE; REIBMAN; TRIVEDI, 1988).

Parametric sensitivity analysis or differential sensitivity analysis is one of the most well-known parametric techniques characterized by a sensitivity index is known as $S_\theta(Y)$, which indicates the impact of a given measure known as Y for a parameter θ. Usually, when dealing with sensitivity analysis techniques, a negative index may indicate that the correspondent parameter negatively impacts this interest metric. In contrast, a positive index indicates that this parameter may improve the metric if it is altered (MATOS et al., 2015).

Equation 2.20 shows how the percentage difference index is calculated for a metric $Y(\theta)$, where $max\{Y(\theta)\}$ and $min\{Y(\theta)\}$ are the maximum and minimum output values, respectively, computed when varying the parameter θ over the range of its *n* possible values of interest.

If $Y(\theta)$ is known to vary monotonically, only the extreme values of θ (i.e., θ_1 and θ_n) may be used to compute $max\{Y(\theta)\}$; $min\{Y(\theta)\}$, and subsequently $S_\theta(Y(\theta))$ (MATOS et al., 2015).

$$\mathrm{S}_\theta(\mathrm{Y}) = \frac{\max\{\mathrm{Y}(\theta)\} - \min\{\mathrm{Y}(\theta)\}}{\max\{\mathrm{Y}(\theta)\}} \tag{2.20}$$

Each $S_\theta(Y)$ is calculated by fixing the other parameters' values. The respective impact is computed for each of the input parameters. Then the most significant effect is found based on rank (MATOS et al., 2015). A higher MTTR may lead to a worse annual downtime, while a higher MTTF may provide a better general availability value.

2.4.3 Stochastic Petri Net

Stochastic Petri Nets (SPN) are a subclass of Petri Nets, used to describe synchronization and concurrency for dynamic discrete event systems, whose dynamic behavior can be represented by Continuous-Time Markov Chains (CTMC) (FLORIN; FRAIZE; NATKIN, 1991; MARSAN et al., 1986). Geometric and exponential distributions define the sensitization durations associated with the transitions in order to be able to build an equivalent Markovian process and, thus, analyze the behavior of the (CARDOSO; VALETTE, 1997) network. In this sense, SPNs are isomorphic CTMCs (MURATA, 1989).

In 1982, M. K. Molloy presented the SPN as a technique capable of specifying systems and presenting a probabilistic analysis of them. They arose from the formalism of Timed Petri Nets (TPN), whose main characteristic is the association of a fixed delay for each transition of the model. Molloy (1982) defined that all SPN transitions were timed and that they had an exponentially distributed delay.

According to German et al. (1995b) a Stochastic Petri Net is defined by the 9-tuple $SPN = \{P, T, I, O, H, \Pi, G, M_0, Atts\}$, where:

- $P = \{p_1, p_2, ..., p_n\}$ is the set of places;
- $T = \{t_1, t_2, ..., t_m\}$ is the set of immediate and timed transitions, $P \cap T = \emptyset$;
- $I \in (\mathbb{N}^n \rightarrow \mathbb{N})^{n \times m}$ is the matrix that represents the entry arcs (which may be dependent on markings);
- $O \in (\mathbb{N}^n \rightarrow \mathbb{N})^{n \times m}$ is the matrix that represents the exit arcs (which may be dependent on markings);
- $H \in (\mathbb{N}^n \rightarrow \mathbb{N})^{n \times m}$ is the matrix that represents the inhibitory arcs (which may be dependent on markings);
- $\Pi \in \mathbb{N}^m$ is a vector that associates the priority level to each transition;
- $G \in (\mathbb{N}^n \rightarrow \{true, false\})^m$ is the vector that associates a guard condition related to marking the place with each transition;
- $M_0 \in \mathbb{N}^n$ is the vector that associates an initial mark of each place (initial state);
- $Atts = (\mathbf{Dist}, \mathbf{Markdep}, \mathbf{Policy}, \mathbf{Concurrency}, W)^m$ comprises the set of attributes associated with transitions, where:
  - $Dist \in \mathbb{N}^m \rightarrow \mathcal{F}$ is a possible probability distribution function associated with the time of a transition (this distribution may be dependent on marking) (the domain of $\mathcal{F}$ is $[0, \infty)$;
  - $Markdep \in \{constante, enabdep\}$, where the probability distribution associated with the time of a transition can be independent ($constant$) or dependent on marking ($enabdep-$ the distribution depends on the current enabling condition);
  - $Policy \in \{prd, prs\}$ defines the memory policy adopted by the transition ($prd-$ emph preemptive repeat different, default value, identical in meaning to emph race enabling policy; $prs-$ emph preemptive resume, corresponds to **age memory policy**);
  - **Concurrency** $\in \{ss, is\}$ is the degree of concurrency of the transitions, where ss represents the semantics **single server** and is represents the semantics **infinity server**.

- $W : T \rightarrow IR^{+} \cup \{0\}$ is the weight function, which represents the $emph(w_t)$ weight of immediate transitions and the $lambda_t$ rate of timed transitions, where:
  $\pi(t) = \begin{cases} \geq 1, & \text{if } t \text{ it's an immediate transition;} \\ 0, & \text{otherwise.} \end{cases}$

  If t is a timed transition, so $lambda_t$ will be the value of the exponential probability density function parameter.

  If t is an immediate transition, so W_t will be a weight, which is used to calculate the probabilities of triggering the immediate transitions in conflicts.

  Inhibitory arcs are used to prevent transitions from being enabled when a specific condition is true.

In SPNs, when multiple transitions are enabled on the same m mark, the transition that has the highest probability of triggering is the transition that has the shortest delay time associated with it (BALBO, 2001). When a SPN transition is triggered, a new mark can be generated in the Petri net formalism. This new mark may contain transitions that were already enabled in the previous mark but have not been triggered. Because of the lack of memory property (*memorelles*) of the exponential distribution, it can be assumed that the activity associated with each transition is resumed for any new markup. The behavior of the variables can be described through a stochastic process, together with the use of *memorelles*. Thus, if an event produces a trigger of the t transition and turns the $M1$ tag into $M2$, the transitions future evolution that was sensitized by $M1$ before the trigger of t should be identical to that transitions would suffer if they were to be sensitized by $M2$. Only the geometric and exponential distributions verify this fact (BALBO, 2001).

Transitions have time, trigger, and transition-delay through random variables distributed exponentially. This time characteristic in the transitions allows different types of operation for the trigger, including the three-phase trigger and the atomic trigger. Three-phase trigger assumes that the marks (*tokens*) are consumed from the place of entry when the transition is enabled and that the marks are produced at the place of exit after the time interval with the transition has elapsed. On the other hand, atomic firing assumes that the marks are consumed from the place of entry and generated at the place of exit only after the transition is triggered (MARSAN et al., 1995).

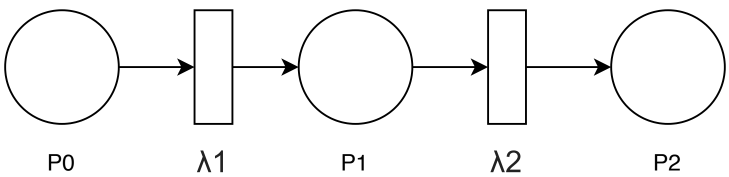
Transitions in SPNs can be immediate or timed. Timed transitions have associated times, which are exponentially distributed. Immediate transitions have associated time equal to zero. In these, the activation period corresponds to the activity's execution period, and the trigger corresponds to the end of the activity. Different levels of priority can be assigned to transitions. The trigger priority of immediate transitions is higher than that of timed transitions. The triggering probabilities associated with immediate transitions can resolve conflict situations (BALBO, 2001; MARSAN et al., 1995).

In SPN models, the transitions are triggered according to the action interleaving semantics (MARSAN et al., 1986). The construction of realistic SPN models serves to evaluate systems and find metrics of dependability and performance. SPN models have two types of states: volatile states (*vanish*) and tangible states (*tangible*). Volatile states are created as a result of marking places that are preconditions for enabling an immediate transition. The term volatile is used because the marks reach these places and are instantly consumed. The permanence time of the markings in these places is zero. Tangible states are created due to marking places that are preconditions for enabling a timed transition (MARSAN et al., 1995). The advantage offered for using stochastic models is the possibility of adding values for specific input parameters in order to obtain output values.

SPN models only consider immediate transitions and timed transitions with exponentially distributed trigger times. These transitions shape actions, activities, and events. A variety of activities can be modeled using the throughput subnets and s-transitions constructors. These constructors represent expolinomial distributions, such as the Erlang, Hypoexponential, and Hyperexponential distributions (DESROCHERS; AL-JAAR, 1995). Combinations of places, exponential transitions, and immediate transitions can be used between two places to represent different types of distributions.

Figure 9 describes a throughput subnet formed by two exponential transitions in series with parameters λ_1 and λ_2, respectively. A mark in place $P0$ will appear in place $P2$ after triggering exponential transitions, which have an associated time $\tau = \tau_1 + \tau_2$, whose density function is given by Equation (2.21) (DESROCHERS; AL-JAAR, 1995).

Figure 9 – Serial Connection with Erlang



Source – Adapted from (DESROCHERS; AL-JAAR, 1995)

$$f_\tau(t) = (f_{\tau 1} * f_{\tau 2})(t) = \frac{\lambda_1\lambda_2(\exp^{-\lambda_1 t} - \exp^{-\lambda_2 t})}{\lambda_2 - \lambda_1}, t \geq 0 \tag{2.21}$$

$*$ is the convolution operator. For the case where λ_1, λ_2 = ... = λ_n, the density function is given by Equation (2.22).

$$f_\tau(t) = \frac{\lambda^n t^{n-1} \exp^{-\lambda t}}{(n-1)!}, t > 0 \tag{2.22}$$

This expression represents an Erlang type distribution of order N. An Erlang type distribution is specified by two parameters $\lambda > 0$ and $n > 0$ (DESROCHERS; AL-JAAR, 1995).

2.5 MAINTENANCE POLICY

The concept of failure implies a break, rupture, discontinuity between expected and verified. Thus, it is a dynamic concept, as it has an underlying action. A defect, in turn, is a physical or functional characteristic, essentially static. However, it is not entirely so, as it can also originate only in the course of the system's functioning itself.

Maintenance strategies can be used to coordinate maintenance activities to each system's particular characteristics, as well as the objectives that are to be achieved, namely minimizing costs, downtime, maximizing reliability, or availability (SOUSA, 2009). This process is called a maintenance policy.

A maintenance policy comprises all actions that must change the system's state to keep it in an operational mode or to return it to an operational condition if it failed (BLISCHKE; MURTHY, 2003). It can be divided into two basic groups *Preventive Maintenance* (PM) and *Corrective Maintenance* (CM).

Preventive Maintenance comprises periodic activities to increase the system's life and reliability. Maintenance may require shutting down a system in operational mode to perform a planned component replacement, testing, and overhaul. PM is crucial to achieving maximum equipment performance, offering the opportunity to detect and correct any problems before they become significant and expensive, thus minimizing the risk of unplanned downtime. On the other hand, corrective maintenance (CM) is performed to restore a failed system to its operational state. The actions involve repairs or replacements (either for new or used items) of all defective parts and components essential for the proper functioning of the entire system (WANG; PHAM, 2006).

2.6 SERVICE LEVEL AGREEMENT

Service Level Agreement (SLA) is a document that defines a set of commitments between the company that provides services and the customer that hired them, where the company must meet the established performance and cost expectations (ALPENDRE, 2006). It specifies services and fees, which must be achieved to fulfill the commitments assumed (STURM; MORRIS; JANDER, 2000).

If used correctly, the SLA should define the customer's needs, provide tools for gauging attributes, simplify complex problems, and eliminate unrealistic expectations (KANDUKURI; PATURI; RAKSHIT, 2009). Additionally, it must present the parties' responsibilities, alternative plans, monitoring reports, and penalties (MCCONNELL; SIEGEL, 2004).

The SLA use depends on two principles: (i) negotiation of conditions under which the services will be provided and must be equally agreed between the parties; (ii) monitoring of these conditions during the use of resources (BENNANI et al., 2015), which can be achieved via tools, which must include means of security and auditing that add reliability.

Each service level in the SLA can be differentiated to define Quality of Service (QoS) parameters. However, SLAs on their own do not have any effect since their value lies in the way they are managed. It is essential to improve the service provider's ability to enforce the contract with the client to minimize the application of penalties (fines) for non-adherence (NASSIF; MARTINS, 2005). The following items show penalty and performance metrics:

- Availability – The service provider is responsible for ensuring the availability agreed in the contract.
- Downtime – The interruption duration must be maintained according to the level of service offered by the contract. The system period in failure mode must be less than the suspension period agreed in the SLA contract.
- Penalty – Considering that the period of inactivity is more extended than the one agreed in the contract, the service provider must pay a fine, which is based on the hours of system interruptions and their cost (ALPENDRE, 2006). These fines are generally based on the quality of the services offered and, when applied, they represent high values (LI; CASALE; ELLAHI, 2010).

The central management challenges are to provide a reliable and adherent service to the SLA, complete monitoring while the service is provided, to correct performance gaps, and to ensure accurate billing based on content and QoS (DÉSOBLIN; PAPINI, 2001).

SLA parameters are identified by a group of metrics, which decide the values to be collected to confirm that the document's terms are being fulfilled (ALJOUMAH et al., 2015). Each service model has its characteristics that differentiate them. Therefore, using metrics that suit specific needs is important. Tables 1, 2 and 3 show metrics for IaaS, PaaS and SaaS, respectively.

Table 1 – Metrics for IaaS

Parameter	Description
CPU	CPU speed for the virtual machine.
Memory	Memory size for the virtual machine.
Initialization	Time it takes the virtual machine to be ready for use.
Storage	Storage size for short- or long-term contract.
Maximum scale	Maximum virtual machines per user.
Minimum scale	Minimum number of virtual machines per user.
Automatic scale	Boolean value for automatic scaling feature.
Increase	Time to increase a specific number of virtual machines.
Decrease	Time to decrease a specific number of virtual machines.
Physical server	Maximum virtual machines that can operate on a single server.
Availability	Service uptime.
Response time	Time taken to receive and complete the process.

Source – Adapted from Aljoumah et al. (2015)

Table 2 – Metrics for PaaS

Parameter	Description
Integration	Integration with systems and platforms.
Scalability	Degree of use with many online users.
Pay for what you use	Charging based on the resources or time of using the service.
Deployment environment	Offline and cloud support.
Browsers	Defines supported browsers, such as: Firefox, Explorer, etc.
Number of developers	Defines how many developers can access the platform simultaneously.

Source – Adapted from Aljoumah et al. (2015)

Table 3 – Metrics for SaaS

Parameter	Description
Reliability	Ability to keep working in most cases.
Usability	Easy to use interfaces.
Scalability	Used by individual or large organizations.
Availability	Service uptime.
Customizable	Flexible to use with different types of users.

Source – Adapted from Aljoumah et al. (2015)

The team must choose the metrics and apply them to generate reports and measure the service's use and provision, ensuring the SLA is effectively complied with, avoiding possible inconvenience.

2.7 DESIGN STANDARDS GUIDE

To solve the problem of the lack of standardization for the implementation of private cloud infrastructures, considering aspects of quality and sustainability, this section intends to offer an approach that presents a set of sustainability guides and indicators to assist in the companies' decision-making process.

The main aspects that the sustainability indicators consider are the social, economic, environmental, and safety aspects. Among the most detailed indexes that can be used within the environmental area are CO_2 emissions, total energy consumed, and the relationship between the weight of waste and product produced. Toxicity information is available for most known chemical compounds and can also be considered as indicators of environmental issues (GODFREY; TODD, 2001).

ISO is a non-governmental organization founded in 1947 in Geneva, and today present in about 189 countries. Its function is to promote the standardization of products and services so that their quality is permanently improved. Standards and norms have significant benefits in demonstrating how best practices can be used. It is possible to point out some results resulting from its implementation, such as: defining the life cycles of the processes, analyzing errors, increasing the degree of safety, organizing working methods, generating continuous improvements, assessing the impacts of change, among others (MAGALHÃES; PINHEIRO, 2007).

It is important to emphasize that the standards do not define the processes to be implemented and, in most cases, they are not mandatory or mandatory and are often adaptable. They present themselves as a facilitator to ensure that project needs are met and are in legal compliance.

Considering all the issues involved and the difficulties faced in the process of planning and implementing private infrastructures, namely due to the lack of information on the part of the companies that build their private data centers - for security reasons, in this section, we look for current standards, techniques and indicators that can guide the first planning moment, especially considering ecological sustainability. In the sequence, we present some standards and techniques, and some environmental sustainability indicators (VALE; ALENCAR, 2020).

2.7.1 ISO IEC 22123 Cloud Computing

This standard defines cloud computing, its terms, and definitions. A version made by ABNT in 2015, in Portuguese, is given by the NBR ISO/IEC 17788 standard (ABNT, 2015). The document establishes a reference for defining cloud computing's scope and application, helping develop national IT markets, serving as support to regulatory agencies. Other efforts to standardize in the cloud computing area include:

- ISO/IEC DIS 19086-2 - Service Level Agreement (SLA).

- ISO/IEC FDIS 19941 - Interoperability and portability.
- ISO/IEC 19944 - Cloud services and devices: data flow, data categories, and data usage.
- ISO/IEC AWI 22123 - Concepts and terminology.
- ISO/AWI 22624 - Taxonomy-based data manipulation for cloud services.
- ISO/IEC NP TR 22678 - Guidance for policy development.
- ISO/IEC 30134-1 and ISO/IEC 30134-2 - Data centers - Main performance indicators.

2.7.2 ITIL

ITIL was formed in the late 1980s by the Central Communications and Telecom Agency (CCTA), now the Office of Government Commerce (OGC), to discipline and allow comparison between the proposals of the various proponents of IT service providers for the British government. During the 1990s, the practices gathered at ITIL started to be adopted by European private organizations, since ITIL was conceived as an open standard, above all due to the significant focus on quality, guaranteed by the definition of processes and the proposition of best practices for IT Service Management, enabling adherence to the ISO 9000 practice and the European Foundation for Quality Management (EFQM) reference model. Among the motivating factors of the current race for the adoption of the practices gathered at ITIL, we can mention the increase in the following aspects (MAGALHÃES; PINHEIRO, 2007):

- Delivery and maintenance costs of IT services.
- Organizational requirements regarding the quality and cost/benefit of IT services.
- Demand to measure the return on IT investments.
- Complexity of IT infrastructure.
- Pace of changes in IT services.
- Need for the availability of IT services.
- Aspects related to security.

ITIL seeks IT services management guidance, describing the objectives, general activities, necessary prerequisites, and expected various processes. Each process consists of interrelated activities, based on a stipulated objective, performed to achieve the desired results. A process can become quite complex, depending on the organization, with a specific management method for each process. The proposed process reference model has two areas in which ITIL processes are fundamental for its full operation (MAGALHÃES; PINHEIRO, 2007):

- Service Delivery.
- Service Level Management.
- Capacity Management.
- Finance Management.
- Support Services.
- Service Desk.
- Incident Management.
- Problem Management.
- Configuration Management.
- Change Management.
- Version Management.

2.7.3 ISO 9000 - Quality Management

ISO 9000 designates a group of technical standards that establish a quality management model for organizations in general, whatever their type or dimension. It states the following: quality can be said when all the characteristics of a product or service required by the customer are delivered to this customer. As far as quality management is concerned, it is stated that it means, then, that the organization guarantees that its products or services satisfy the customer's requirements and that they are in compliance with any regulations applicable to those products or services (ABNT, 2015; VHP, 2009).

Quality management requires a business understanding of quality and value and assurance that the service will be designed and managed to meet these specifications for the IT services department. Total Quality Management continually encourages everyone to meet internal and external customers' demands to achieve a competitive advantage. It is a generic term used to describe a vast collection of philosophies, concepts, methods, and tools (ABNT, 2015; VHP, 2009).

Willian E. Deming introduced to a diagram the methodology known as Plan-Do-Check-Act (PDCA), presenting the processes of planning and implementing services, which consists of four basic tasks (VHP, 2009):

- Plan - establishes the objectives and processes necessary to deliver quality services.
- Do - implements the processes established in the plan.

- Check - monitors and establishes metrics for the processes to confirm that they are being executed with quality.
- Act - takes actions to improve processes and the results generated by them.

Both ISO 9000 and ISO/IEC 20000 (VHP, 2009) include this cycle in their continuous improvement approach.

2.7.4 ISO 20000 - IT Service Management

This standard does not formalize the inclusion of ITIL practices, although a set of management processes is described that is aligned with the processes defined within ITIL. ISO 20000 defines requirements for the correct management of a company providing IT services, guaranteeing customers the delivery of quality services. The requirements are the definition of policies, objectives, procedures, and management processes to ensure sufficient quality in IT services provision (VHP, 2009).

The first activity of service delivery processes elaborates service level agreements between the requesting areas and the IT service management area. Like the ITIL delivery service, the service delivery processes in ISO 20000 also deal with issuing and reporting service availability and continuity, budget and cost accounting, and capacity management. Besides, customer needs are identified, and change management for those needs. Customer feedback should also be obtained by measuring their satisfaction level (MAGALHÃES; PINHEIRO, 2007).

IT Service Management (ITSM) manages all processes that cooperate to ensure the quality of IT services in production, according to the service levels agreed with the customer. The main benefit of ITSM is that it provides quality criteria for end-to-end customer-focused services. This is the only basis for the mature management of the IT infrastructure, represented by IT components and non-IT, grouped into services in operation or at other service life stages (VHP, 2009).

2.7.5 ISO 14000 - Environmental Management

ISO 14000 (ABNT, 2005a) dictates rules for an Environmental Management System (EMS) to promote socio-economic balance in the face of current needs, thus providing organizations the assistance to achieve their environmental and economic objectives. This standard is also based on the PDCA methodology. A technical committee was created and subdivided into several others to develop the 14000 series standards, responsible for:

- Subcommittee 1 - Standards on environmental management systems (EMS).
- Subcommittee 2 - Standards related to environmental audits.
- Subcommittee 3 - Rules on environmental labeling.

- Subcommittee 4 - Environmental performance standards.
- Subcommittee 5 - Life cycle analysis standards.
- Subcommittee 6 - Rules on definitions and concepts.
- Subcommittee 7 - Rules on the integration of environmental aspects in product design and development.
- Subcommittee 8 - Environmental Communication Standards.
- Subcommittee 9 - Rules on climate change.

Of the standards created, one of the best known is ISO 14001, it does not establish specific environmental performance criteria, but it does apply to any organization that wants to implement an EMS and ensure compliance with its environmental policy (ABNT, 2005b). Other well-known standards are:

- ISO 14004 – EMS for the internal part of the company.
- ISO 14010 - Environmental audit.
- ISO 14031 - Performance of the EMS.
- ISO 14020 - Product labeling and environmental declarations.

The EMS helps companies control environmental issues holistically, enabling them to develop and implement a policy that considers the legal requirements on significant environmental aspects. Many organizations have carried out analyzes to assess their environmental performance; however, such analyzes may not be sufficient to provide an organization with meeting the legal requirements and its policy (VALLE, 2002).

Systematically, the adoption and implementation of a set of environmental management techniques can help obtain optimal results for all interested parties to be appropriate and economically viable. The cost-benefit ratio of such techniques is taken fully into account (ABNT, 2005a).

2.7.6 ANSI/TIA-942

The primary purpose of ANSI/EIA/TIA-942 – Telecommunications infrastructure standard for data centers – is to provide regulations for planning DC installations, with rules for structured cabling, standardized architectures, protocol instructions, media recommendations and distance between distribution paths, horizontal cabling, tier guidance, topologies, temperature and humidity requirements, energy efficiency, among others (DIMINICO; JEW, 2006) (see Subsection 2.1.2).

The standards are presented to guide the metrics and techniques that should be considered to define better private cloud infrastructures (IaaS) or data centers.

2.8 FINAL CONSIDERATIONS

This chapter introduced cloud computing and data centers, their classifications types, architectures, and systems. The concepts for dependability and sustainability were also presented. The formalism of modeling techniques used in this work, such as EFM, RBD, and SPN, were showed. Subsequently, the definitions of maintenance policy and SLAs were given. Finally, it was also presented a set of techniques to notify companies and designers to use more efficient and sustainable practices in the elaboration phase of the architecture projects. The techniques are found in several pieces of literature; however, there is a certain granularity that makes it difficult to find more specific guides, as organized in this chapter.

3 METHODOLOGY

This chapter presents the methodology of this research. Initially, the methodological choices are presented, and in the sequence, the phases that led to the study's conduct are specified. Lastly, the tools used are shown.

3.1 METHODOLOGICAL CHOICES

Scientific research enables an approximation and an understanding of the reality to be investigated, the process of which is permanently unfinished. We reach successive approximations of reality through it, which provides us with subsidies for an intervention in the real (GERHARDT, 2009).

LEHFELD e Barros (1991 apud GERHARDT, 2009) refers to research as being the inquisition, the systematic and intensive procedure, which aims to discover and interpret the facts that are inserted in a given reality.

The methodological procedures aim to identify the processes and activities carried out to study and solve a given problem (SANTOS, 2018). We carry out this research with the methodological choices summarized in Table 4 and detailed in sequence.

Table 4 – Summary of Methodological Choices

Concerning	**Classification**
Approach	Quantitative and qualitative
Nature	Basic
Objectives	Exploratory
Procedures	Bibliographic Research (Systematic Review)
Method	Hypothetical-deductive

Source – Elaborated by the author.

This research is classified according to definitions given in Gerhardt (2009), as described in the sequence:

- As for the approach, as quantitative and qualitative, we carried out a meta-analysis on the studies related to this research's context while conducting a systematic literature review.
- As for nature, primary research was initially carried out to generate knowledge to solve specific problems involving local truths and interests.
- As for the objectives, the research is classified as exploratory, given the intention of becoming familiar with the problem, making it more explicit. We carried out a bibliographic

review and a systematic literature review to support the problem, the models, and the experiments.

- As for the procedures, there is a bibliographic search, which shows the systematic review results. We carried out the experiments in the models' validation phase, aiming to create scenarios for observing and manipulating variables to measure their effects on the study object.

- We have chosen the method as the hypothetical-deductive, considering that the research starts with a problem or a gap in scientific knowledge, passing through the formulation of a research hypothesis and a deductive inference process, which tests the prediction of the occurrence of phenomena covered by that hypothesis.

3.2 RESEARCH METHODOLOGY

The methodology used in this work's conception considers some steps presented in a generic way to describe the activities to obtain the results. The steps for carrying out this research are shown in Figure 10.

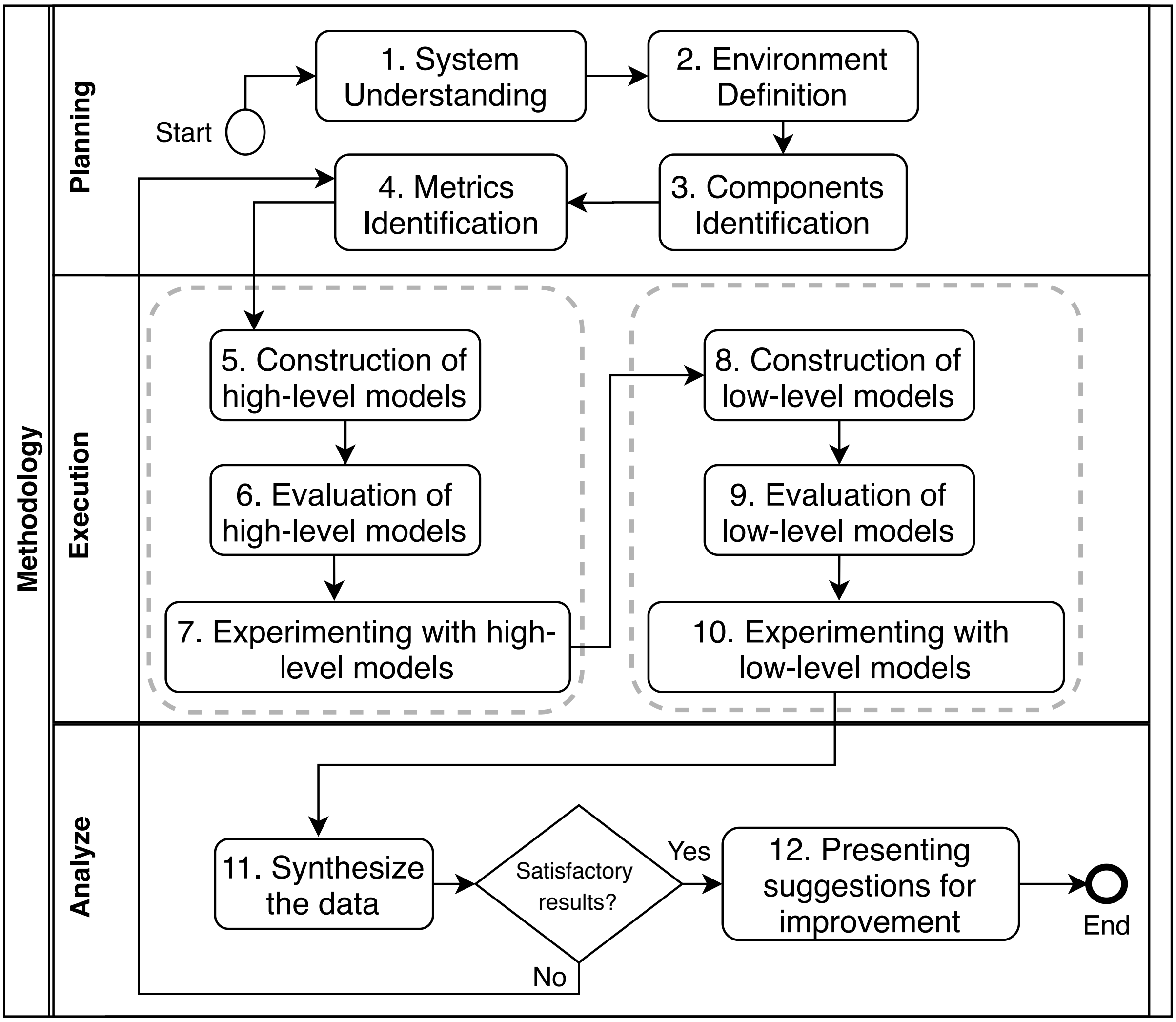
Initially, we classified three phases for conducting the specified activities, as described next:

- **Planning** - In this phase, we carry out the activities that correspond to the system's general understanding to define which architectures and components to be represented and identify and find the metrics that serve as input parameters of the models.

- **Execution** - In this stage, we perform the modeling, which is done in a hybrid and hierarchical way to design the EFM, RBD, and SPN models. These models represent four data center power and IT architectures, which have different redundancy levels. Thus, we created four EFM models for the energy architectures and their respective RBD models (another four). We created seven RBD models for representing networks, ranging from the high-level representation of a blade server's components to a tier 4 DC network. Finally, we created an SPN (low-level) model, representing a maintenance approach for data center networks and verifying if, with the parameters defined in SLA classes, it is possible to achieve the minimum availability suggested for each tier (considering the components of the physical architectures).

  For each type of model, we carry out experiments. For EFMs models, we create scenarios based on a load of data centers tiers 1 to 4 to represent the use of different types of raw materials used in energy production and thus quantify CO_2 emissions. We create scripts for RBDs models (both the RBDs of the electrical and network architecture) to perform the sensitivity analysis to identify the components' influence on systems' availability. For the SPN model, we defined nine SLA classes, varying the intervals between carrying out

Figure 10 – Phases and Activities of this Research.



Source – Elaborated by the author.

preventive maintenance in three classes and varying the mean time to repair (MTTR) for each class. In this way, we capture the availability achieved by each SLA class, and at the end, we compare whether it was possible to achieve the minimum availability suggested for each tier, according to the maintenance policy used.

- **Analysis** - In this phase, we did the data synthesis to understand and identify whether the results correspond to the research objectives and then present suggestions to improve sustainability, availability, reliability, and maintainability.

The idea of presenting the steps sequentially and individually is to abstract the entire lengthy research process, which continues until reaching the results. We implemented general activities at each stage to describe a set of products and prepare the next stage's environment assets. Some activities can be performed interactively. However, in most of them, the process was sequential. The activities that comprise the research stages, listed in Figure 10, are as follows:

1. System Understanding – In this activity, it is essential to identify the study area and the problem to be analyzed. We identify the research problem, the central research question, and five research questions that guided the SLR, which was performed to obtain a more significant theoretical basis and find evidence of gaps for research and the problems open in the study area and the relevance of the problem defined for this study. The conduct of the systematic review is extensively detailed in Chapter 4. However, it is vital to clarify that the process was fully systematized following well-established literature guides, being conducted according to a previously prepared protocol, and assisted in using a specific tool (StArt).

2. Environment Definition – In this activity, we understand the environment's characteristics, such as architectures and minimum components necessary to represent the models better. We follow very well defined guides in the literature, such as *Uptime Institute* and ANSI/TIA/EIA-942 standard (presented in Sections 2.1.1 and 2.1.2), to get to know and become familiar with the environment to be modeled. Given the studied definitions, we decided to design the models with minimum redundancy of each tier's physical components. In this way, it was easier to modularize and fit the models by tier.

3. Components Identification – This activity represents the understanding of the characteristics and connections between the architectures' components. Specifically, for electrical architectures, there was a more significant concern in choosing the most efficient components, given the idea of proposing energy savings. An example is that we opted for the ATS component instead of the STS.

4. Metrics Identification - This step aims to define the metrics, which will be considered in the experiments' evaluation process. These metrics are variables capable of influence the expected results. Thus, after each metric's value definition, it is fundamental to observe the experiment results to identify positively or negatively the results.

5. Construction of high-level models – Corresponds to the EFM and RBD models' construction to represent the electrical architectures and data center networks, according to component redundancy specifications by tier. We consider a high level of abstraction (therefore, we call them high-level models) because the modeling, as mentioned earlier techniques, represents only the architectures' physical characteristics. Abstracting the logical characteristics, there is an excellent reduction in possible ways of modeling the desired architectures (unlike state-based models for logical representations). Besides, because the literary specifications are already well documented for the redundancy levels, it was possible to reduce the complexity in understanding the system and modeling. We perform modeling in a hybrid and hierarchical way. Hybrid, due to the use of three types of different modeling techniques. Hierarchical, due to the results of primary models being used as input parameters for subsequent models.

6. Evaluation of high-level models – After performing the modeling, we can perform the evaluations searching for interest metrics. We expect sustainability, availability, and reliability aspects such as energy efficiency, impacts on the environment, and operating costs, among others, will be defined.

7. Experimenting with high-level models – To validate the results, we present some scenarios. In this work, the specific objective is to demonstrate how individual raw material sources can negatively impact sustainability due to CO_2 emissions into the atmosphere. We also show the importance of the components for the system's availability through a sensitivity analysis for all RBD models of electrical and network architectures.

8. Construction of the low-level model - Corresponds to the proposed SPN model, which allows the representation of a maintenance policy, in several states to characterize the need for preventive and corrective repairs. We consider a low level of abstraction, given the possibility of having dependencies between states, according to the needs of the maintenance policy intended to be represented. Thus, the state space with the SPN modeling technique can be huge, presenting a much higher complexity level (compared to the two EFM and RBD modeling techniques), requiring the states' explicit representation and storage. With the SPN technique, we define how preventive maintenance will be triggered so that a team performs repairs on the architecture components to avoid errors in the system's functioning. We also defined how to restore the system after serious failures have occurred, in which case corrective maintenance is triggered and performed by the same preventive maintenance team.

9. Evaluation of the low-level model – Aims to provide the estimated results (e.g., availability and downtime) for the proposed maintenance policy. For each tier (1 to 4), we use the MTTF resulting from the evaluation of the RBD model of the networks in the states that represent the system failure in the SPN model, so we vary the SLA classes to observe how the defined policy can influence the availability of the system. Thus, it is possible to define whether the SLA classes are represented to meet the tiers' needs (availability suggested in the literature). For the techniques based on simulation, e.g., SPN, the simulation is engaging because it does not need the infrastructure's previous implementation to evaluate its functioning and propose improvements. Adjustments are possible to optimize results, both in the input metrics and in the states' representation.

10. Experimenting with the ow-level model – We define the parameters to carried out the experiments after the models' metrics have been evaluated. These assessments aim to find results that may lead to conclusions and suggestions for improvements to the analyzed systems. In our SPN, the results must compose the SLAs' attributes to verify the services achieved from the proposed maintenance approach are satisfactory to tiers' needs.

11. Synthesize the data – We check and synthesize the results to identify possible improvements in the metrics, models, or experimentation scenarios.

12. Presenting suggestions for improvement – After carrying out all the activities mentioned, we can finally identify the represented architectures' biggest bottlenecks and then suggest improvement points according to the study objects (components, metrics, and connections).

Although the steps presented to define this methodology are generic, in this work, the process was accomplished by adopting a hybrid modeling strategy, which offers combinatorial models and models based on states to represent the chosen architectures and the maintenance policy for data center networks. Although the characteristics are defined in a particular way for the proposed infrastructures, we can manipulate the attributes and adapt the techniques to obtain a new representation if necessary.

3.3 SUPPORT TOOLS

The main tools used to perform this work were StArt to support conducting a systematic review, and Mercury to execute the modeling and evaluation. Both are briefly described next and are more detailed in Appendices A and B.

3.3.1 StArt

The systematic literature review (SLR) has many stages and activities that can be performed laboriously and repetitively. Therefore, the support of a computational tool is essential to improve the quality of the application. The tool called StArt (State of the Art through Systematic Review) aims to assist the researcher, supporting this technique's application (START, 2020). For this work, we use the graphical interface in version 3.0.3. For more details see Appendix A.

3.3.2 Mercury Tool

We use the graphical interfaces for EFM, RBD, and SPN in version 4.8 of the Mercury Tool (MACIEL et al., 2017) for this work. Mercury is software for supporting performance, dependability, and energy flow modeling quickly and powerfully. The software provides graphical interfaces for creating and evaluating SPNs, Continuous-Time Markov Chains (CTMCs), Discrete-Time Markov Chains (DTMCs), RBDs, Fault Trees (FTs), and EFMs (MACIEL et al., 2017). We also use the scripting language to perform dependability evaluation and sensitivity analysis. Mercury Scripting Language was designed to allow greater flexibility for evaluating models, extracting metrics, and generating graphs and reports. Additionally, it is possible to carry out different experiments, whose hierarchical relationships between models can be supported, in contrast to the graphical interface (MACIEL et al., 2017). For more details see Appendix B.

3.4 FINAL CONSIDERATIONS

In this chapter, the methodological choices and methodology of this research were addressed. The methodological choices comprise the approach, nature, objectives, procedures, and method of the research. Furthermore, the methodology corresponds to the phases and activities included in the development of this work. Three phases of the methodology were defined: planning, execution, and analysis, containing 12 general activities. A brief description of the main tools used was also provided.

4 SYSTEMATIC LITERATURE REVIEW

This chapter presents the Systematic Literature Review (SLR), which was fulfilled to obtain a diagnosis of the area and the problem studied. The procedures performed to conduct the systematic protocol, map the evidence, and consolidate the results are explained throughout the sections. A description of the works that most correlate with our research is also presented, most of them having been achieved from RSL. They are indexed in two sections, according to the topics of interest.

4.1 CONTEXTUALIZATION

To understand the process of conducting an SLR, we presented some essential concepts. First, it is essential to distinguish between bibliographic review (or literature review) and SLR. The first is conducted in a traditional way in which no rule or protocol is followed; in the second, the whole process is accomplished in a systematic way in which if it is inferred that if others follow the same steps, the same results must be found (at least until the selection phase of the studies in the bases).

According to Kitchenham e Charters (2007), experimental studies consist of testing a hypothesis and may include quantitative and qualitative analysis, which can be classified into primary, secondary, and tertiary studies. Primary studies investigate a specific research question and can be executed through experiments, case studies, or survey research. Secondary studies review primary studies related to one or some research questions with the specific objective of integrating/synthesizing evidence related to that research questions. For that, a systematic literature mapping or a systematic literature review can be performed.

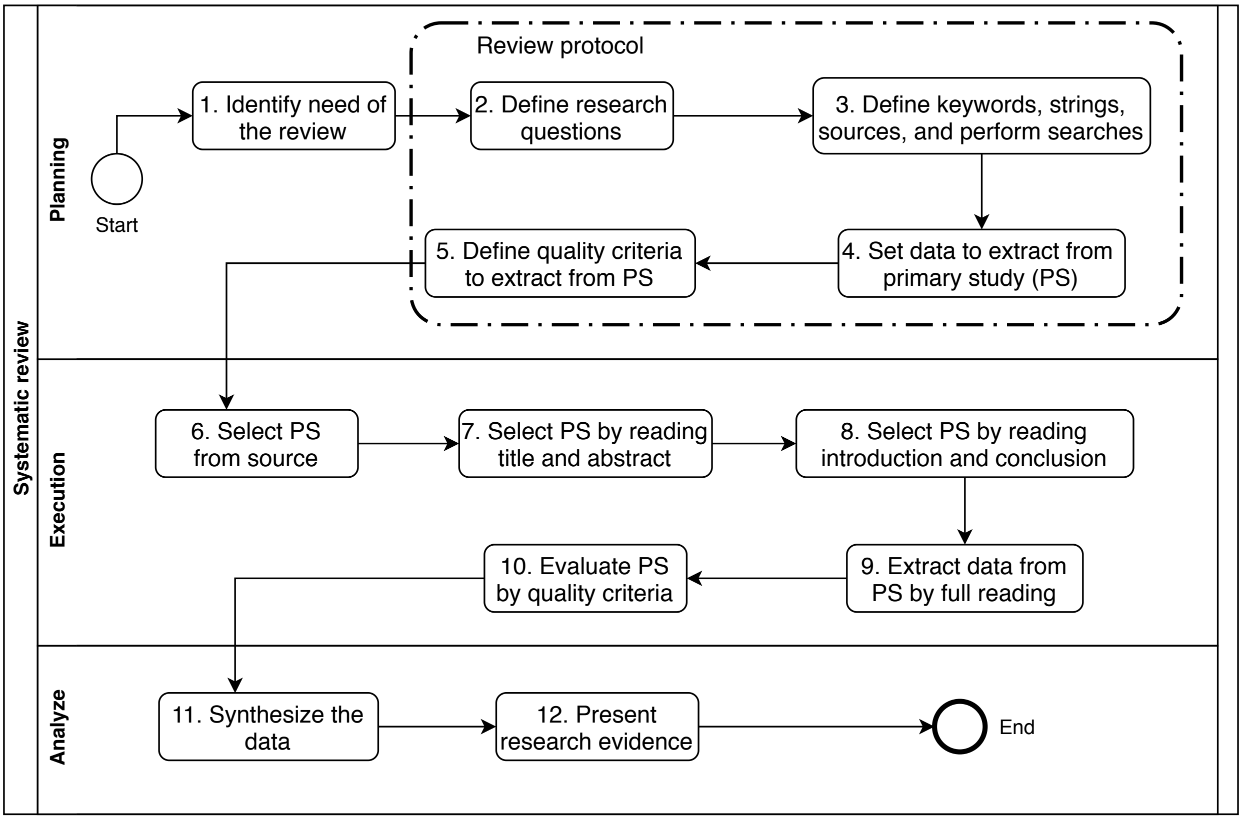
A systematic mapping starts from the performance of pre-shaped research for a systematic review, but, during the research, the low relevance of results is observed, characterizing a granularity concerning the possible results and scope in the research, aiming to provide a view wider area of research, and determine whether there is the research evidence on a topic, indicating the amount of evidence (KITCHENHAM; CHARTERS, 2007). To perform a systematic mapping, an entire mapping protocol must be followed. After the systematic mapping is completed, then a systematic review should begin.

An SLR is a means of identifying, evaluating, and interpreting all available research relevant to a particular research question, topic area, or phenomenon of interest. Individual studies that contribute to a systematic review are called primary studies (KITCHENHAM, 2004).

Several reasons justify conducting a systematic review. Briefly, it can be said that it is possible to summarize evidence on a certain area, identify gaps for further investigation, and provide a structure for new research activities. The systematic review's conduct must be implemented fairly, following a pre-defined research strategy, which allows assessing the research's integrity and identifying which ones support or not their hypotheses. (KITCHENHAM, 2004).

A series of discrete activities are conducted at different stages of the process to implement an SLR. The phases must obey a sequential order of execution. However, the activities started during the development of the protocol can be refined at a later stage. It is essential to clarify that the phases must be followed in sequential order: planning, execution, and analysis. However, selecting inclusion or exclusion criteria and selecting the priority level of articles can serve as examples of tasks that can be performed when the researcher deems them appropriate. Figure 11 shows the phases of this process and their respective activities.

Figure 11 – Phases for Conducting a Systematic Literature Review



Source – Elaborated by the author.

In the planning phase, the research questions must be formulated to search for primary studies. In the execution phase, the necessary data must be extracted to answer the research question. In the analysis process, the data must be synthesized so that the questions can be answered. It is essential to differentiate between the systematic mapping and systematic review. Both follow the same protocol for conducting the research. However, the mapping characteristics are more quantitative, broad, and generally summarized in graphs. In contrast, the SLR characteristics are more qualitative, with descriptive critical analyzes, thorough and in-depth, to elucidate new shreds of evidence and relevant aspects by summarizing the selected studies. For this thesis, we present data from a systematic literature review.

4.2 PLANNING

An SLR protocol was created at this stage, which consisted of defining the work's objectives, problematization, research hypotheses, research questions, keywords and search strings, search bases, and inclusion and exclusion criteria. The protocol follows the recommendations given in (KITCHENHAM, 2004; KITCHENHAM; CHARTERS, 2007) and the activities are described in sequence.

4.2.1 Research Questions

The SLR presented was performed in two moments. The first to delimit gaps and open problems (three research questions were created), and the second to identify the state of the art of related works (two more research questions were elaborated). Initially, the research area was defined, and some problems were raised; interest has always been focused on cloud computing, IaaS, data center, reliability, non-functional requirements (NFR), and sustainability. Then, we search for a context whose solution is better related to this thesis's results to identify gaps and solutions that could enter as contributions of this research. To guide the direction of the study, a central research question was initially created, based on a research hypothesis. Then, five research questions (RQ) were defined, in a more extended way, which, when answered, could enrich and strengthen the theoretical knowledge about the problem domain, the gaps, and list the related works. The central research question, the hypothesis that guided the search, and the other questions created are presented next.

Research Question: How is it possible to identify bottlenecks in data center architectures to assist in the design phase, aiming to provide greater sustainability when dependability requirements are also met?

Hypothesis: The hierarchical and hybrid modeling approach makes it possible to find architectural bottlenecks to suggest improvements before implementing data centers.

To answer the central question, confirm the research hypothesis, and list related works, we try to answer the research questions presented in Table 5.

Table 5 – Research Questions

#	Research Questions
RQ1	Is there an approach, technique, method, methodology, model, or a guide to support the migration to cloud computing?
RQ2	Is there an approach, technique, method, methodology, model, or a guide to support migration to cloud computing, which considers non-functional requirements and sustainability?
RQ3	Which are the open problems related to the research area?
RQ4	Which are the most important dependability requirements (attributes) for sustainable data center projects? (What are the biggest concerns?)
RQ5	Which models/modeling techniques are most used for proposing sustainable data center infrastructures?

Source – Elaborated by the author.

The questions are descriptive and classifying. In the initial stage of research, these types of questions are asked in the exploration phase, when the researcher tries to understand the phenomena and identify useful distinctions to clarify the understanding.

4.2.2 Search Strategies

The keywords and their synonyms were defined to compose the research questions' search strategies (strings). The strings use the logical operators OR to perform the search for any of the presented synonyms and the AND operator for adding more words to the term. Several pilot tests were performed to calibrate the strings. Table 6 presents the search strategies for each research question.

Table 6 – Search Strings

RQ	Search Strings
RQ1	("cloud computing" OR "private cloud" OR "public cloud") AND ("approach" OR "technique" OR "method" OR "guide" OR "models") AND ("migration")
RQ2	("cloud computing" OR "private cloud" OR "public cloud") AND ("approach" OR "technique" OR "method" OR "guide" OR "models") AND ("migration") AND ("sustainability" OR "sustainable") AND ("non-functional requirements")
RQ3	("cloud computing" OR "private cloud" OR "public cloud") AND ("approach" OR "technique" OR "method" OR "guide" OR "models") AND ("migration") AND ("open issues" OR "load mapping" OR "open research" OR "open problem")
RQ4	("cloud computing" OR "private cloud" OR "public cloud" OR "data center" OR "datacenter" OR "data centre" OR "IaaS" OR "infrastructure-as-a-service") AND ("implantation" OR "project" OR "migration") AND ("non-functional requirements" OR "attributes" OR "dependability") AND ("sustainability" OR "sustainable" OR "energy efficiency")
RQ5	("data center" OR "datacenter" OR "data centre" OR "IaaS" OR "infrastructure-as-a-service") AND ("model" OR "modeling" OR "simulation model" OR "simulation modeling" OR "analitic model" OR "analitic modeling") AND ("evaluation" OR "requirement evaluation" OR "performance evaluation" OR "dependability evaluation" OR "performability evaluation" OR "sustainability evaluation" OR "analysis" OR "requirement analysis" OR "performance analysis" OR "dependability analysis" OR "performability analysis" OR "sustainability analysis") AND ("sustainability" OR "sustainable" OR "energy efficiency") AND ("energy flow model" OR "reliability block diagram" OR "stochastic petri nets" OR "markov chain")

Source – Elaborated by the author.

4.2.3 Search Sources

We selected the search sources considering their importance for the research area and the number of results achieved in the pilot tests, choosing those that presented the best results. All searches were performed electronically, using the mechanisms offered by the websites of the chosen databases. It is worth mentioning that some provide better resources for inserting terms, offering several filters that allow us to select articles within a specific date range, by author, by keywords contained in the title only, in the abstract or text. For the searches of this research, we used no filter. A specific date range was not considered, aiming to reach all articles that presented the keywords cited since the technology's appearance. Another point to note is that the chosen words could be within any part of the text. The search bases chosen are shown in Table 7.

Table 7 – Search Bases

Search Source	URL
IEEE Xplorer	http://www.ieeexplore.ieee.org/Xplore
ACM Digital Library	http://dl.acm.org/
Scopus	http://www.scopus.com

Source – Elaborated by the author.

4.2.4 Inclusion and Exclusion Criteria

The inclusion and exclusion criteria should identify the primary studies that provide direct evidence on the research issue. They must be piloted to ensure that they can be interpreted reliably and classify the studies correctly (KITCHENHAM; CHARTERS, 2007). Tables 8 and 9 present the inclusion and exclusion criteria for this research, respectively.

Table 8 – Inclusion Criteria

#	Inclusion Criteria
1	The studies must have been published in journals, symposia, or conferences of the mentioned bases.
2	The studies must be written in English.
3	The studies must be available on the web.
4	The studies must be complete articles.
5	Studies that address relationships with the focus of work.
6	Primary studies.

Source – Elaborated by the author.

Table 9 – Exclusion Criteria

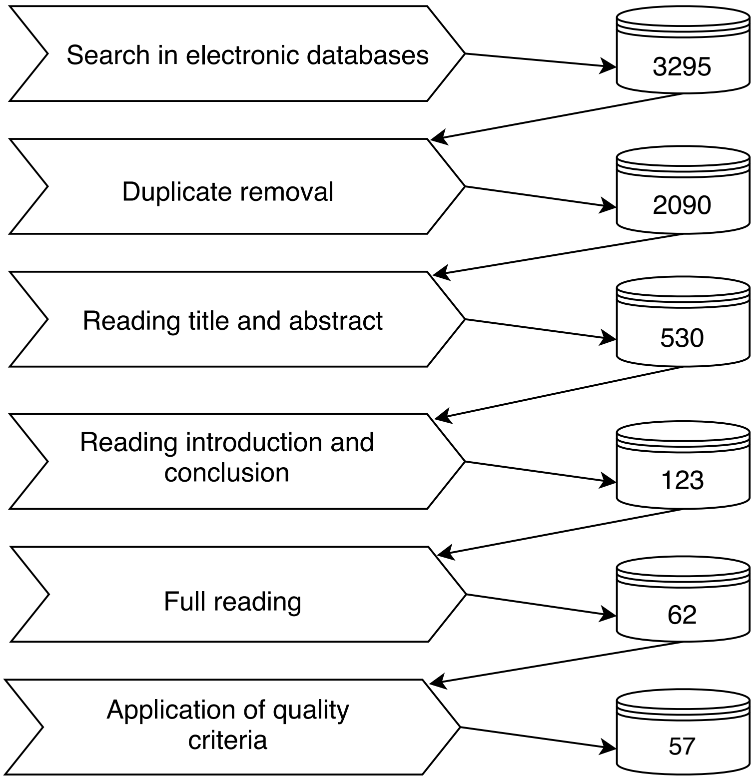
#	Exclusion Criteria
1	Secondary and tertiary studies.
2	Summarized articles.
3	Duplicate studies (the only one will be considered).
4	Studies that do not address the relationship with the focus of the work.
5	The redundant study by an author (the complete version will be considered).
6	Book or book chapter.

Source – Elaborated by the author.

4.3 EXECUTION

The execution phase comprises selecting the studies and extracting the data. With the StArt tool's aid, the *.bib* of the selected studies for each research question were imported from the three search bases mentioned. The process for conducting the protocol started in 2016, and it was necessary to update the searches in 2017 and 2020. In total, 3295 studies went through six stages of selection, as shown in Figure 12.

Figure 12 – Articles Selection



Source – Elaborated by the author.

After reading the full articles, the inclusion and exclusion criteria defined in the research protocol were applied, and we recorded the results for further analysis. In the end, we read in full a total of 123 articles, of which 62 answered at least one of the research questions and remained for the quality analysis phase. After the application of the quality criteria, remain 57 articles for data collection.

4.4 ANALYSIS

In this analysis phase, we seek to synthesize and interpret the results to find evidence that the studies meet the purpose of the systematic review, that is, that it answers at least one of the research questions. As shown in Figure 12, some quality criteria were applied, which resulted in the rejection of four articles. The quality evaluation criteria are used to classify the selected studies by a technical score capable of measuring their credibility, completeness, and relevance. This score represents one point when the answer is yes (Y = 1), zero when the answer is no (N = 0) and is worth half a point when you answer partially (P = 0.5). We classify the studies, an average of the responses was made, which corresponds to the quality index achieved (this is given in percentage – shown in Table 15), articles with an average below seven (7.0) points were rejected (which happened with the four articles previously mentioned). To calculate the quality index, Equation 4.1 was used. The quality criteria for this research are shown in Table 10.

$$Quality_{index} = (\frac{Sum_{responses}}{14}) \times 100 \tag{4.1}$$

Table 10 – Criteria for Quality Assessment

#	Criteria	Responses
1	Is there an explanation of why the study was done?	Y=1, N=0, P=0.5
2	Was the study based on research (or is it based on the author's experience)?	Y=1, N=0
3	Do the authors make clear what the purpose of the study is?	Y=1, N=0, P=0.5
4	Is the proposed approach clearly described?	Y=1, N=0, P=0.5
5	Is the research context clearly described (laboratory, used products)?	Y=1, N=0, P=0.5
6	Has the research context been described at an appropriate level (industry, laboratory environment, products used and so on)?	Y=1, N=0, P=0.5
7	Is there a discussion about the results obtained?	Y=1, N=0, P=0.5
8	Are the limitations of the study clearly described?	Y=1, N=0, P=0.5
9	Is there a clear presentation of the open problems in the study area?	Y=1, N=0, P=0.5
10	Is there enough information for the study to be replicated?	Y=1, N=0, P=0.5
11	Is the study supported by a tool?	Y=1, N=0
12	If the article deals with NFR, have they been validated?	Y=1, N=0, NA
13	If the article is about NFR, is it also about the phases of the requirements engineering process? If so, what are they?	Elicitation, analysis and negotiation, specification, management, NA
14	Can the research also add value to the industrial community?	Y=1, N=0, P=0.5

Source – Elaborated by the author.

An important decision is that criteria 12 and 13 were only considered for articles that dealt with non-functional requirements because this research is interested in three different themes. Thus, studies that had not addressed NFR would not be unfairly scored. The primary studies (PS) that reached a quality index above 70% after applying the criteria then went on to the data extraction phase. These are presented in Table 11 with their respective identifier, publication year, title, and reference. The data collected and organized to show general and specific piece of information about the selected studies are defined in Table 12. To answer questions 16 to 19 of this table, the results are presented in Section 4.5 (Table 23), considering that the questions refer to understanding the current state of the art. It was also interesting to check if the articles presented new definitions for terms related to this research and if they answered attributes related to the research questions. The attribute is used only as a reference for question extensions. These are presented in Table 13.

Table 11 – Selected Primary Studies

Id	Year	Title	Reference
PS_01	2010	Efficient resource management for Cloud computing environments	(YOUNGE et al., 2010b)
PS_02	2013	Cloud Migration for SMEs in a Service Oriented Approach	(NUSSBAUMER; LIU, 2013)
PS_03	2014	Enhancing energy efficiency in resource allocation for real-time cloud services	(BAGHERI; ZAMANIFAR, 2014)
PS_04	2014	An Approach of Quality of Service Assurance for Enterprise Cloud Computing (QoSAECC)	(CHU et al., 2014)
PS_05	2013	Return on Security Investment for Cloud Platforms	(TSALIS; THEOHARIDOU; GRITZALIS, 2013)
PS_06	2015	Enterprise-scale cloud migration orchestrator	(HWANG et al., 2015)
PS_07	2014	Heterogeneity-Aware Workload Placement and Migration in Distributed Sustainable Datacenters	(CHENG; JIANG; ZHOU, 2014)
PS_08	2015	Evaluating a cloud federation ecosystem to reduce carbon footprint by moving computational resources	(GIACOBBE et al., 2015)
PS_09	2014	Overlay energy circle formation for cloud data centers with renewable energy futures contracts	(EROL-KANTARCI; MOUFTAH, 2014)
PS_10	2012	A business-driven framework for evaluating cloud computing	(SRIPANIDKULCHAI; SUJICHANTARARAT, 2012)
PS_11	2013	Robustness and oppurtuneness based approach for Cloud deployment model selection	(SONI; NAMJOSHI; PILLAI, 2013)
PS_12	2015	Cloud migration: Planning guidelines and execution framework	(CHANG; CHIU; CHIAO, 2015)
PS_13	2013	Effort Estimation in Cloud Migration Process	(SUN; LI, 2013)
PS_14	2013	Decision Support in Data Centers for Sustainability	(PAWLISH; VARDE; ROBILA, 2013)
PS_15	2014	Building Green Cloud Services at Low Cost	(BERRAL et al., 2014)
PS_16	2014	A framework architecture based model for cloud computing adaptive migration	(ABDERRAHIM; CHOUKAIR, 2014)
PS_17	2011	Simulator improvements to validate the Green Cloud Computing approach	(WERNER et al., 2011)
PS_18	2015	The migration of the university IT infrastructure toward a secure IaaS Cloud	(MOHAMED; KARIM; AHMED, 2015)
PS_19	2014	Why Do Companies Migrate Towards Cloud Enterprise Systems? A Post-Implementation Perspective	(BOILLAT; LEGNER, 2014)
PS_20	2015	Assurance of Security and Privacy Requirements for Cloud Deployment Model	(ISLAM et al., 2015)
PS_21	2010	Enabling Sustainable Clouds via Environmentally Opportunistic Computing	(WITKOWSKI et al., 2010)
PS_22	2013	Optimizing water efficiency in distributed data centers	(REN, 2013)
PS_23	2013	A survey on energy and power consumption models for Greener Cloud	(PRIYA; PILLI; JOSHI, 2013)
PS_24	2009	Data analysis, visualization and knowledge discovery in sustainable data centers	(MARWAH et al., 2009)
PS_25	2012	Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing	(BELOGLAZOV; ABAWAJY; BUYYA, 2012)

Id	Year	Title	Reference
PS_26	2016	A Review of the Current Level of Support to Aid Decisions for Migrating to Cloud Computing	(ALKHALIL; SAHANDI; JOHN, 2016)
PS_27	2017	Determinants of adoption of cloud computing services by small, medium and large companies	(PALOS-SÁNCHEZ; ARENAS-MÁRQUEZ; AGUAYO-CAMACHO, 2017)
PS_28	2017	Cost-benefit analysis-evaluation model of cloud computing deployment for use in companies	(MARESOVA; SOBESLAV; KREJCAR, 2017)
PS_29	2017	Consideration of marginal electricity in real-time minimization of distributed data centre emissions	(DANDRES et al., 2017)
PS_30	2018	A colored generalized stochastic Petri net simulation model for service reliability evaluation of active-active cloud data center based on IT infrastructure	(LIU et al., 2017)
PS_31	2011	A Formal Approach to the Quantification of Sustainability and Dependability Metrics on Data Center Infrastructures	(CALLOU et al., 2011)
PS_32	2019	A methodology to assess the availability of next-generation data centers	(ROSENDO et al., 2019)
PS_33	2020	A model-based strategy for quantifying the impact of availability on the energy flow of data centers	(VALENTIM; CALLOU, 2020)
PS_34	2015	A modeling approach for cloud infrastructure planning considering dependability and cost requirements	(SOUSA et al., 2014)
PS_35	2020	An algorithm to optimise the energy distribution of data centre electrical infrastructures	(FERREIRA et al., 2020)
PS_36	2013	An algorithm to optimize electrical flows	(FERREIRA et al., 2013)
PS_37	2015	An Algorithm to Optimize Electrical Flows of Private Cloud Infrastructures	(FERREIRA et al., 2015)
PS_38	2019	An artificial neural network approach to forecast the environmental impact of data centers	(FERREIRA et al., 2019)
PS_39	2014	An integrated modeling approach to evaluate and optimize data center sustainability, dependability and cost	(CALLOU et al., 2014)
PS_40	2017	Cloud infrastructure planning considering different redundancy mechanisms	(SOUSA et al., 2017)
PS_41	2014	Dependability evaluation of data center power infrastructures considering substation switching operations	(SILVAA et al., 2014)
PS_42	2018	Design of fuel cell powered data centers for sufficient reliability and availability	(RITCHIE; BROUWER, 2018)
PS_43	2011	Estimating reliability importance and total cost of acquisition for data center power infrastructures	(FIGUEIRÊDO et al., 2011)
PS_44	2013	Estimating sustainability impact of high dependable data centers: A comparative study between Brazilian and US energy mixes	(CALLOU et al., 2013b)
PS_45	2018	Estimating the environmental impact of data centers	(FERREIRA et al., 2018)
PS_46	2020	Evaluating the impact of maintenance policies associated to SLA contracts on the dependability of data centers electrical infrastructures	(MELO; JUNIOR; CALLOU, 2020)

Id	Year	Title	Reference
PS_47	2020	Formal models for safety and performance analysis of a data center system	(BENNACEUR; KLOUL, 2020)
PS_48	2016	Geographical Load Balancing across Green Datacenters: A Mean Field Analysis	(NEGLIA; SERENO; BIANCHI, 2016)
PS_49	2010	Impact analysis of maintenance policies on data center power infrastructure	(CALLOU et al., 2010)
PS_50	2012	Integrated reliability modeling for data center infrastructures: A case study	(MÜLLER; STRUNZ, 2012)
PS_51	2018	PLDA-D an algorihm to reduce data center energy consumption	(FERREIRA et al., 2018)
PS_52	2014	Power Availability Provisioning in Large Data Centers	(SANKAR; GAUTHIER; GURUMURTHI, 2014)
PS_53	2010	Quantifying the Sustainability Impact of Data Center Availability	(MARWAH et al., 2010)
PS_54	2019	Reliability and availability evaluation for cloud data center networks using hierarchical models	(NGUYEN et al., 2019)
PS_55	2019	Stochastic models for optimizing availability, cost and sustainability of data center power architectures through genetic algorithm	(AUSTREGÉSILO; CALLOU, 2019)
PS_56	2011	Sustainability and dependability evaluation on data center architectures	(CALLOU et al., 2011)
PS_57	2012	Towards a Net-Zero Data Center	(BANERJEE et al., 2012)

Source – Elaborated by the author.

Table 12 – Data Extraction Form

#	Study Data	Description
1	Study identification	Unique Id
2	Author, year, title, country, extraction date, .bib	General information
3	Number of Citations	Identify how cited the study is
4	Means of publication	Newspaper, Conference, Symposium
5	Application context	Industrial, Academic
6	Research method	Controlled experiment, case study, survey, ethnography, action research, illustrative scenario, not applicable
7	Research maturity	Validation research, evaluation research, solution proposal, philosophical articles, opinion articles, experience articles, not applicable
8	Work's objective	Identify the purpose of the work
9	Are there definitions for cloud computing?	Identify concepts of interest
10	Are there definitions for sustainable cloud computing?	Identify concepts of interest
11	Are there definitions of RNF for cloud computing?	Identify concepts of interest
12	Are there definitions of sustainable RNF for cloud computing?	Identify concepts of interest
13	Attributes related to RQ1	Cloud migration support
14	Attributes related to RQ2	RNF and sustainability for the cloud
15	Attributes related to RQ3	Identify open problems
16	Architecture modeled	Electrical, cooling, IT
17	What approaches?	–
18	Models Type ?	RBD, EFM, SPN, MC, CTMC, GSPN, CPN, Other
19	Analysis Type	Sustainability, dependability, sensitive, maintainability

Source – Elaborated by the author.

Table 13 – Attributes of Research Questions

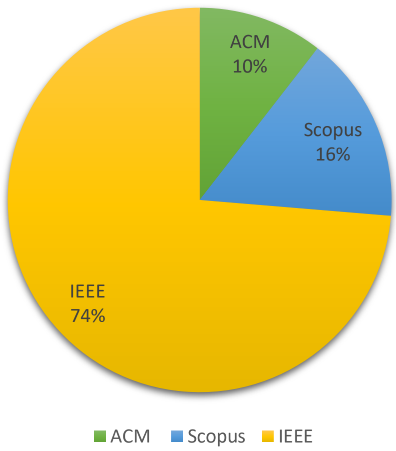
RQ	Attributes	Responses
RQ1	Is there an approach, technique, method, methodology, model, guide to support the migration to cloud computing?	Yes, no
	If so, what is it?	Approach, technique, methodology, method, guide, none
	Specifies for which type of cloud?	Yes, no
	If so, which one?	Public, private, hybrid, community cloud
	Is there a specific service model?	Yes, no
	If so, which one?	SaaS, PaaS, IaaS, DaaS, DbaaS, CaaS, EaaS, others, several
	If specifies other or several service models, which are they?	–
RQ2	Are NFR specified?	Yes, no
	If so, what are they?	–
	Have the NFRs been validated?	Yes, no
	If there is a validation technique, what is it?	–
	Considers sustainability?	Yes, no
	If so, does it address any specific pillar?	Yes, no
	If so, what is it?	Social, economic, environmental
	Which aspects considered?	Energy efficiency, alternative energies, environmental concerns, refrigeration, water efficiency, water footprint
RQ3	Does it present open problems or future works?	Yes, no
	If so, which are they?	–
	Does it have limitations?	Yes, no
	If so, which are they?	–

Source – Elaborated by the author.

4.4.1 Analysis Results

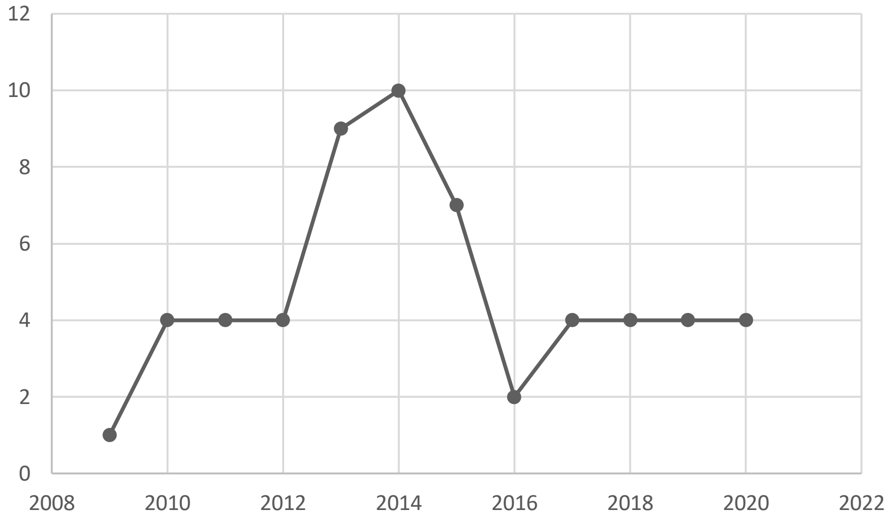
Figure 13 presents the percentage of studies selected for data extraction by the search base. By the classification results, it can be said that the choice of the IEEE database was significant, given that the studies that answered the research questions better are in most of this database. The four studies rejected by the quality index were also on this basis but did not correspond to the questions designed to assess studies' quality, considering this research's context. A study was also consummated on the year of publications, illustrated in Figure 14.

Figure 13 – Articles per Base



Source – Elaborated by the author.

Figure 14 – Articles per Year



Source – Elaborated by the author.

In a detailed analysis of the study area, nine more focus on migration with gaps for future work or open problems. One study kept the focus on non-functional requirements for the cloud, two also maintained this focus but also considering possible future work. Four studies address sustainability without focusing on cloud migration precisely but specifying cloud computing topics in some way. Another nine studies deal with sustainability showing future projects and one mix of migration, sustainability, and problems.

What can be concluded from the presentation of this data is that only three articles answer three research questions simultaneously, none of which specifically consider non-functional requirements for sustainability in the cloud. This information is essential evidence that much remains to be done and opens a gap for further studies on this particular topic, which may generate new research questions for future mapping.

Once exposed that no article addressed non-functional requirements for sustainability in the cloud, it can be said that there are no answers to research question 2 (RQ2) in its original form. Therefore, there was a need to make an adaptation, which resulted in two distinct issues, one to consider articles that address migration to cloud and non-functional requirements and another to consider articles that address migration to cloud and sustainability, thus generating the research questions (RQ2.1 and RQ2.2), which are:

- RQ2.1 Is there an approach, technique, method, methodology, model, a guide to support migration to cloud computing that considers **non-functional requirements**?
- RQ2.2 Is there an approach, technique, method, methodology, model, a guide to support migration to cloud computing that considers **sustainability**?

However, this fact had already been noticed since the strings' calibration during the pilot tests' performance, where some bases did not show results when placing the complete string created for RQ2. Likewise, strings were also adapted when necessary, which made it possible to achieve good search results.

Overall, 16 articles can answer RQ1. That is, they address the topic of migration to cloud computing; only three studies respond to RQ2.1; a total of 14 articles respond to RQ2.2; 21 studies respond to RQ3. Based on these data, we can conclude that most articles have satisfactory answers to this search. Considering that many proposals for future work have been presented and can be seen as gaps to be studied, we concluded that there are initiatives to solve problems about migration to the cloud and that sustainability is on the rise, but little is consolidated about these technologies or patterns. Very little has been achieved about NFR, which can represent disinterest in the topic or be considered under some other name that was not described in the search string. Moreover, 28 articles respond to RQ4 and RQ5.

4.4.2 Evidence Mapping

At this moment, we present the evidence obtained through this systematic review, in which the results that classify the articles to the research questions RQ1, RQ2.1, RQ2.2, and RQ3. For articles that answer questions RQ4 and RQ5, data are presented only in Section 4.5.

As already mentioned, 16 studies respond to RQ1; only three respond to RQ2.1, 14 respond to RQ2.2, and 21 respond to RQ3. Table 14 presents primary studies versus research questions. This table facilitates quick identification if the study answers one, two, or three of the research questions (none answers all).

Table 14 – Primary Studies X Research Questions

Id	RQ1	RQ2.1	RQ2.2	RQ3
PS_01			X	X
PS_02	X	X		
PS_03			X	
PS_04	X			X
PS_05	X			X
PS_06	X			
PS_07			X	X
PS_08			X	
PS_09			X	X
PS_10	X			X
PS_11	X			X
PS_12	X			X
PS_13	X			X
PS_14			X	X
PS_15			X	
PS_16	X			
PS_17	X		X	X
PS_18	X			
PS_19	X			X
PS_20	X	X		X
PS_21			X	X
PS_22			X	X
PS_23			X	X
PS_24			X	
PS_25			X	X
PS_26	X	X		X
PS_27	X			X
PS_28	X			X
PS_29			X	X

Source – Elaborated by the author.

It is worth mentioning that no study addresses NFR and sustainability at the same time. Therefore, there are no studies that answer the four questions simultaneously. Only two articles answer three research questions, they are: PS_17 (WERNER et al., 2011) and PS_26 (ALKHALIL; SAHANDI; JOHN, 2016). However, the PS_17 answers to RQ1, **RQ2.2**, and RQ3, while the PS_26 answers to RQ1, **RQ2.1**, and RQ3. Thus, PS_17 addresses cloud migration, sustainability, and open issues, and PS_26 addresses cloud migration, non-functional requirements, and open issues. We can consider that these two articles are the most complete in terms of questions that can answer and have a high relationship with this research's focus. However, unlike both, this study aims to address non-functional requirements (some treated here as dependability attributes - see Section 2.2) and sustainability together.

The indexes of the quality criteria reached by the articles selected for this SLR are shown in Table 15 (Calculated from the Equation 4.1). The articles that answered the three RQs, mentioned in the previous section, PS_17, and PS_26, reached indexes of 95.5% and 81.4%, respectively. Considering that the PS_17 reached the highest quality index presented (together with the PS_14), we can conclude that this preliminary study has substantial evidence to contribute a lot to this research. The articles' indexes fell because the limitations, tools, future work, or description of the research context were not usually presented.

Table 15 – Quality Index

Índex(%)	**Id**
72,9%	PS_02
75%	PS_08, PS_24
81,4%	PS_26
83,3%	PS_03, PS_06, PS_07, PS_18, PS_22, PS_23
84,3%	PS_20
86,4%	PS_16
87,5%	PS_11, PS_12, PS_13, PS_19
90%	PS_27, PS_28
90,9%	PS_15
91,7%	PS_01, PS_04, PS_05, PS_10, PS_21, PS_25, PS_29
95,5%	PS_14, PS_17

Source – Elaborated by the author.

Table 16 shows which articles were selected in each of the search bases. Despite the low participation of ACM, its only study, namely PS_26, was one of those that answered three research questions, together with PS_17, which was indexed by the IEEE database.

Table 16 – Eletronic Databases

Fonte	**Id**
ACM	PS_26
SCOPUS	PS_22, PS_23, PS_24, PS_25, PS_27, PS_28, PS_29
IEEE	PS_01, PS_02, PS_03, PS_04, PS_05, PS_06, PS_07, PS_09, PS_10, PS_11, PS_12, PS_13, PS_14, PS_15, PS_16, PS_17, PS_18, PS_19, PS_20, PS_21

Source – Elaborated by the author.

The PS_25 was the most cited study of all selected, with 1,404 citations, which obtained a quality index of 91.7% and was published in 2012 by SCOPUS. For this research, it answers two questions related to sustainability and open problems. The second most cited study is PS_01, with 237 citations, a quality index of 91.7%, and was published in 2010 by IEEE. For this research, it answers two questions related to sustainability and open problems. These data lead us to believe that there has been a massive interest in sustainability since 2010.

Of the 29 articles that respond to questions RQ1 to RQ3, only 19 of them have definitions for *cloud computing*. Table 17 shows what they are. The RQ1 of this SLR sought to identify an approach, technique, method, methodology, model, or guide to support cloud computing migration. Table 18 presents studies that answer this research question. Considering the data presented in the tables above, in which 19 studies define cloud and 17 studies respond to RQ1, we can see that two studies present definitions for cloud but do not support migration. These are PS_03 and PS_29, whose focus is both on sustainability for the cloud, answering RQ2.2.

Of the 17 articles that support migration to the cloud, 14 of them present an approach, 1 study presents a methodology, and 2 presents a migration technique. Only four articles specify which type of cloud they are considering, the studies being PS_05, PS_06, and PS_18 for a private cloud and PS_10 for a public cloud. A total of seven articles specifies which type of model they support, 4 of them (PS_05, PS_10, PS_13, and PS_18) for IaaS, 2 of them (PS_19 and PS_20) for SaaS, and 1 (PS_04) says that it fits for several models (IaaS, PaaS, SaaS, DaaS).

Table 17 – Articles Featuring Definitions for Cloud Computing.

Id
PS_01, PS_02, PS_03, PS_04, PS_10, PS_11, PS_12, PS_13, PS_16, PS_18, PS_19, PS_20, PS_21, PS_23, PS_25, PS_26, PS_27, PS_28, PS_29

Source – Elaborated by the author.

Table 18 – Articles that Support Migration to Cloud Computing.

Id
PS_02, PS_04, PS_05, PS_06, PS_10, PS_11, PS_12, PS_13, PS_16, PS_17, PS_18, PS_19, PS_20, PS_26, PS_27, PS_28

Source – Elaborated by the author.

Only three articles answer RQ2.1. That is, they specify non-functional requirements, being the studies: PS_02, PS_20, and PS_26. In the PS_02 study, the requirements mentioned are cost, confidentiality, flexibility, performance, security, service, and support. In the PS_20 study, the requirements mentioned are mostly focused on security. They are access control, availability, integrity, and privacy. In the PS_26 study, the requirements mentioned are agility, compliance, cost, performance, availability, guarantee, integrity, privacy, security, and usability. None of the articles presents any technique for validating NFRs, and none considers or specifies what would be sustainable non-functional requirements for the cloud. These studies make little contribution to this project's objective, considering that the context in which they were applied is not related to the desired objectives. The NFRs they cite are well known in the literature.

On sustainability for *cloud computing*, only 13 articles present definitions. Table 19 shows what they are. The RQ2.2 of this SLR sought to identify whether the study considers sustainability. Table 20 presents studies that answer this research question. Considering the data presented in the tables above, in which 13 studies define sustainability and a total of 14 responses to RQ2.2, it is clear that a study that answers this question has no definitions for sustainability. This is PS_07. In general, no definition presented by the other articles stands out from those already widely addressed in the literature. Therefore, no new concept has been identified. Of the 14 articles dealing with sustainability, only 3 (PS_14, PS_15, and PS_21) identify a specific pillar, which is the environmental one. The rest do not present any specific information about this definition. Although the majority did not highlight this point, they all identified which aspects of sustainability would be considered. Table 21 presents what these aspects are, showing how many and which studies refer to them.

Table 19 – Articles that Define Sustainability for Cloud Computing.

Id
PS_01, PS_03, PS_08, PSPS_09, PS_14, PS_17, PS_21, PS_22, PS_23, PS_24, PS_25, PS_27, PS_29

Source – Elaborated by the author.

Table 20 – Articles Considering Sustainability.

Id
PS_01, PS_03, PS_07, PS_08, PS_09, PS_14, PS_15, PS_17, PS_21, PS_22, PS_23, PS_24, PS_25, PS_29

Source – Elaborated by the author.

Table 21 – Sustainability Aspects.

Aspects	Amount	Id
Energy efficiency	3	PS_01, PS_03, PS_08
Energy efficiency, Alternative energies	1	PS_07
Energy efficiency, Environmental concerns	4	PS_17, PS_23, PS_25, PS_29
Energy efficiency, Environmental concerns, Alternative energies	3	PS_09, PS_14, PS_15
Energy efficiency, Cooling, Environmental concerns	1	PS_21
Energy efficiency, Cooling, Environmental concerns, Alternative energies	1	PS_24
Water efficiency, Water footprint, Environmental concerns	1	PS_22

Source – Elaborated by the author.

From the data presented, we define that for all articles that address sustainability, the pillar is considered explicitly as the environmental one. However, it is still necessary to analyze how these aspects influence the other two pillars (social and economic), given that when energy efficiency is achieved in a certain way, there is an impact on cost reduction because the three are intrinsically related.

Following the reasoning of the conclusions reached in each of these subsections, in which only the studies PS_05, PS_06, and PS_18 consider a private cloud infrastructure but do not focus on NFRs. Articles that consider NFRs do not focus on the private cloud, and articles that consider sustainability do not focus on the private cloud or NFRs. We concluded that the objects presented for this research are presented as original and unpublished.

To identify gaps for possible problems, this SLR introduced RQ3. Table 22 shows which articles answer this question.

Table 22 – Articles with Open Problems

Id
PS_01, PS_04, PS_05, PS_07, PS_09, PS_10, PS_11, PS_12, PS_13, PS_14, PS_17, PS_19, PS_20, PS_21, PS_22, PS_23, PS_25, PS_26, PS_27, PS_28, PS_29

Source – Elaborated by the author.

In total, 21 articles addressed, in some way, possible gaps that can be better investigated and appear as an open problem in the area. Numerous problems have been identified, among which are most related to this research, are the following:

- Migration to the cloud - among the most cited problems are inefficient management, problems with the legislation, lack of standards, lack of control, transparency, and licensing.
- Difficulties in meeting quality or financial requirements for the cloud - among the most cited are the security aspects (such as the vulnerability of virtualization, data security, privacy, and confidentiality), performance, availability, and costs.
- Sustainability - lack of practical application; culture plastering; inefficiency of inspection and collection; difficulties in the decision-making process; diversity and high level of complexity of techniques to assess aspects such as energy efficiency; rapid modification of techniques and the emergence of new needs; high cost of more efficient equipment and delay in return on investment.

In addition to the problems mentioned, some opportunities were also identified that deserve further investigation. They are:

- to explore a scheduling system that is both energy conscious and thermal to maximize energy savings;
- to reduce energy costs in a cloud federation ecosystem;
- to integrate energy storage techniques for sustainable computing in green data centers;
- to evaluate the performance and electricity bills of operators in a green data center;
- to measure data center energy usage to predict factors like PUE and find trends in energy efficiency data;
- to investigate NFRs like security and fault tolerance to devise specific metrics to evaluate those NFRs in cloud applications.

Given the gaps presented, in this thesis, we try to contend with some of them, as mentioned in the objectives.

4.5 RELATED WORKS

Much work has been done to address gaps and issues related to cloud computing, infrastructure as a service, and data center in recent years. Researchers have used different approaches to deal with the various problems inherent in distributed, parallel, and on-demand computing, with increasingly urgent requirements given the high availability and reliability charged to suppliers. The problems of migration, implementation, performance, availability, sustainability, and security related to these areas are very abundant.

Our research presents two central lines: sustainability and energy efficiency in data centers and dependability assessment. In recent years, exclusively due to the significant problem of global warming, the scientific community has made a great effort to present solutions to the industrial community to improve technological capacities, mainly concerning natural ecosystems. For this cause, our work and many others are presented as a contribution to mitigating the side effects of overuse of IT, such as the emissions of CO_2 in the atmosphere.

A second systematic literature review was carried out to identify the works related to this research, which was presented as a single SLR for simplification purposes. 922 papers were selected on three search bases. During the selection phase, 850 were rejected, and 72 were accepted. In the extraction phase, 43 were denied, 1 was duplicated, and 28 were accepted. In the end, 28 papers remained that have a strong connection to this research. Among the problems most cited by related works are:

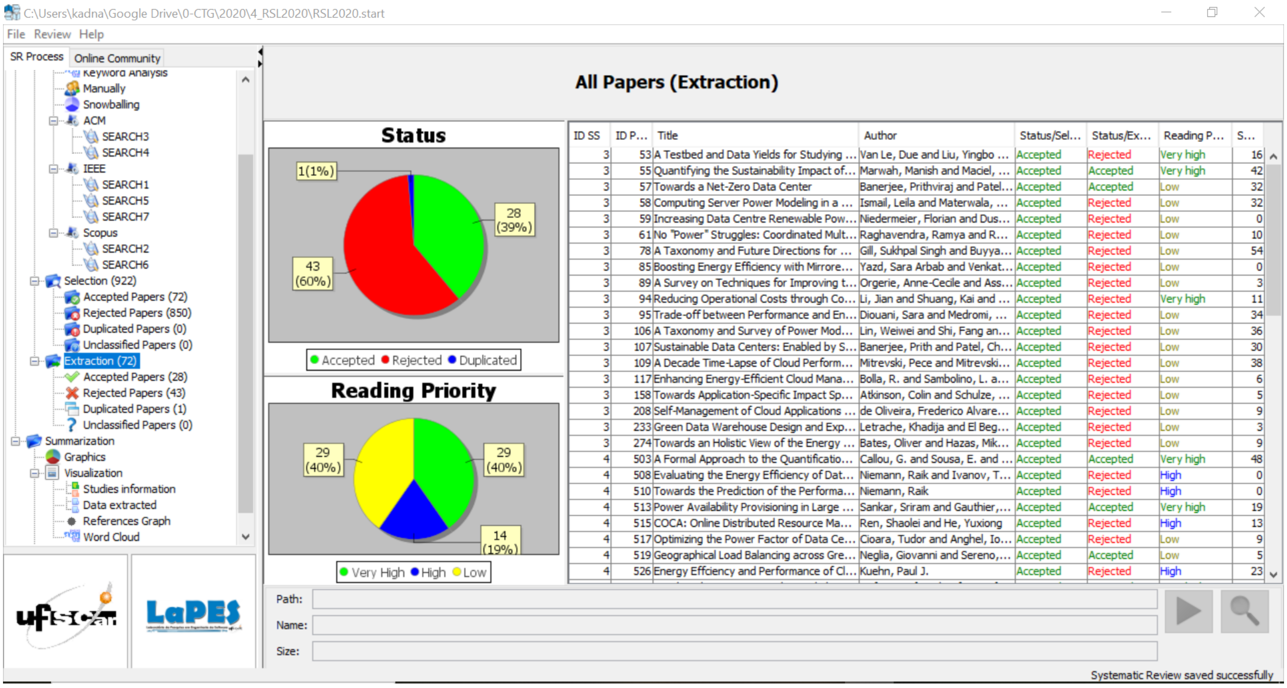
- Data center design - complexity of the energy, cooling, and network infrastructures; high energy consumption; inefficient cooling; lack of physical space; security risks, costs with physical spaces and maintenance, great diversity of technical specialization, interconnections between subsystems, difficulties in meeting performance requirements and high availability.

Considering all the identified gaps, we try to propose a solution capable of mitigating some of these problems.

Figure 15 shows the StArt interface with the percentage of accepted, rejected, repeated articles (first graphic) and the priority ranking related to the first similarity identification (second graphic). For more details, see Appendix A.

Next, we present the works that have a more significant correlation with this research. We divided it into two sections to better modularize the research topics. Note that we chose the one considered complete to be presented in the Table 24 for some works or set of works by the same author. Moreover, we made some differentiations right away for some works that have a more remarkable similarity to our proposal. Some works were added manually, because the language could be in Portuguese or because it is a thesis or dissertation or not being indexed in the chosen databases.

Figure 15 – StArt Tool - SLR for Related Works.



Source – StArt Tool (START, 2020).

4.5.1 Research on Sustainability and Energy Efficiency

Júnior (2019) proposes strategies to improve efficiency and estimate the energy consumption of data centers. An algorithm for distributing electric current in-depth (PLDA-D) is proposed to optimize energy distribution in electrical infrastructures. The modeling includes EFM, SPN, RBD, and CTMC models. The author also uses a multilayer artificial neural network (with the perceptron), predicting the data center's energy consumption based on historic. Other works by the same author appeared as extensions of his research. In Ferreira et al. (2018), the authors carry out electrical architecture evaluations of a tier 3 data center to estimate costs and CO_2 emissions. They also present an energy matrix according to the energy costs from different primary sources for countries like China, Germany, the United States, and Brazil. Additionally, an artificial neural network is used to forecast energy consumption for the following months. Ferreira et al. (2013), Ferreira et al. (2015), Ferreira et al. (2018), and Ferreira et al. (2020) propose and present a power load distribution algorithm in-depth search (PLDA-D) to optimize private cloud electrical infrastructures' power distribution. The PLDA-D adopts the Energy Flow Model (EFM) as the basis for designing power infrastructures. Different scenarios presented the effectiveness of the algorithm. Ferreira et al. (2019) used a multi-layered artificial neural network, which could forecast consumption over the following months, based on the data center's energy consumption historic. All these features were supported by a tool, the applicability of which was demonstrated through a case study that computed the CO_2 emissions and operational costs of a data center using the energy mix adopted in Brazil, China, Germany, and the USA.

- Our research differs from that mentioned in several points. Next, we mention the main ones:
    1. We present the costs and sustainability metrics for four DC electrical architectures, ranging from tier 1 to 4, where our proposal goes from the preparation of the IT demand calculation, according to the number of components to be represented in the network models, to the choice of the most efficient and dense components to compose the access layer of the data centers, as is the case of Blade System. In this sense, we try to make our modeling strategy as close to reality as possible.
    2. We use different materials to quantify CO_2 emissions in contrast to those presented.
    3. We present RBDs models to perform a sensitivity analysis of the electrical and IT architectures. This objective was not presented by the work mentioned.

Callou (2013) proposes a set of models for the integrated quantification of the sustainability impact, cost, and dependability of data center power and cooling infrastructures. The approach taken to perform the system dependability evaluation employs a hybrid modeling strategy that recognizes EFM, RBD, and SPN's advantages. Besides that, a model is proposed to verify that the energy flow does not exceed the maximum power capacity that each component can provide (considering electrical devices) or extract (assuming cooling equipment). Other works by this author remain focused on assessments of sustainability. Callou et al. (2013a) consider a methodology to support the modeling, evaluation, and optimization processes in data center energy systems using Mercury tool for the construction of RBD, SPN, and EFM. Callou et al. (2011) present a set of formal models for estimating sustainability impact and dependability metrics, supported by an integrated environment, namely, ASTRO. Callou, Andrade e Ferreira (2019) propose a set of models for estimating availability, cost, and sustainability. A case study is conducted to illustrate the proposed models' applicability for analyzing IT data center systems. Callou et al. (2014) propose an integrated approach to estimate and optimize high availability levels demanded against the low sustainability impact and cost values with the support of the developed environment, Mercury. Callou et al. (2013b) propose a set of models for the integrated quantification of the sustainability impact, cost, and dependability of data center power and cooling infrastructures. Valentim e Callou (2020) used the technique of modeling colored Petri nets (CPN), responsible for quantifying the cost, environmental impact, and availability of the data centers' electricity infrastructure. Such proposed models are supported by the developed tool, where data center designers do not need to know CPN to compute the metrics of interest. A case study was proposed to show the applicability of the proposed strategy. Melo, Junior e Callou (2020) adopt SPN to evaluate the impact of maintenance policies on the data center electrical infrastructures and present case studies comparing different SLAs. Austregésilo e Callou (2019) propose a multi-objective optimization approach based on Genetic Algorithms, to optimize cost, sustainability, and availability of data center power infrastructures, to maximize availability and minimize cost and exergy consumed. The authors adopt EFM, RBD, and SPN. Two case studies are conducted to show the applicability of the proposed strategy: (i) considers five typical data center architectures that were optimized to conduct the validation process of the proposed strategy; (ii) uses the optimization strategy in two architectures classified by ANSI/TIA-942 (TIER I and II).

- Our research differs from this by representing the modeled architectures and all three points cited for the first set of works mentioned before.

Figueirêdo (2011) proposes new indices to quantify the importance of components, relating costs, which can assist data center designers. Additionally, the author quantifies the impact on the sustainability of data center infrastructures. Improvements are also proposed in the RBD and SPN modules of the Mercury Tool (formerly ASTRO). Figueirêdo et al. (2011) propose an approach that combines reliability importance indexes with TCA (Total Cost of Acquisition) to evaluate data center power infrastructures. The proposed approach aims at indicating the benefits and drawbacks of different architectures for power infrastructures to guide the decision-maker in evaluating of the most appropriate design.

- The authors present different architectures, but they are not classified considering the redundancy levels as classified for the tiers. So, our research differs from this by representing the modeled architectures and all three points cited for the first set of works mentioned before.

Souza (2013) proposes models to contemplate the effect of temperature variation in the data center infrastructures. It is also proposed to assist in elaborating and evaluating different scenarios to find the effect of temperature variation on the data center's IT infrastructures' availability.

Silvaa et al. (2014) propose a methodology, which includes a hierarchical heterogeneous modeling technique that considers the advantages of both SPN and RBD to evaluate data center power infrastructures considering substation switching operations.

Marwah et al. (2010) present an approach to estimate the sustainability impact of data centers. Availability is computed using Stochastic Petri Net (SPN) models, while an exergy-based life-cycle assessment (LCA) approach quantifies sustainability impact. The approach is demonstrated in real-life data center power infrastructure architectures.

Banerjee et al. (2012) present a blueprint for a "net-zero data center": one that offsets any electricity used from the grid via adequate on-site power generation that gets fed back to the grid at a later time. The authors discuss how such a data center addresses the total cost of ownership, illustrating that contrary to the oft-held view of sustainability as "paying more to be green", sustainable data centers—built on a framework that focuses on integrating supply and demand management from end-to-end — can concurrently lead to lowest cost and lowest environmental impact.

Sankar, Gauthier e Gurumurthi (2014) propose a provisioning methodology for the power delivery infrastructure called power availability provisioning that addresses this challenge and provide observations on power infrastructure design based on industry experience operating large data centers.

Müller e Strunz (2012) present a method of calculating the reliability of integrated data center infrastructure models (including the power grid and cooling systems) using Boolean algebra.

Ritchie e Brouwer (2018) propose a design alternative to use multiple fuel cell systems, each supporting a small number of servers to eliminate backup power equipment provided the fuel cell design has sufficient reliability and availability. Potential system designs are explored for the entire data center and individual fuel cells. RBD analysis of the fuel cell systems was accomplished to understand the systems' reliability without repair or redundant technologies.

Bennaceur e Kloul (2020) investigate the different interactions between the data center subsystems (electrical, thermal, and network) and their impact on the data center reliability, and they use Production Trees, a new modeling methodology for dealing with availability issues of production systems.

Younge et al. (2010a) presented a new structure that provides efficient ecological improvements within a scalable cloud computing architecture, using energy-conscious scheduling techniques, variable resource management, live migration, and a minimal virtual machine design.

Reis (2009) made an investigation of green aspects in the implementation of a data center in the industrial area of Suape-PE, where the decisions taken to achieve a higher degree of sustainability are presented.

Albuquerque (2013) presented a set of metrics, models, and tools that assist in assessing both the environmental and operational performance of supply chains. For the evaluation of the supply chain, parameters such as the means of transport used are considered, the distribution strategies adopted, the relationship between producers, intermediaries, and consumers, and the inventory replenishment strategy. High-level modeling is done using *stochastic reward net* (SRN), based on Petri nets, to carry out the experiments. In addition to traditional performance metrics, the proposed models also include indicators for assessing environmental performance, such as energy and resources consumed, exergy, and global warming potential (GWP).

Although the works cited, have characteristics related to this research's objective, such as promoting environmental sustainability and using techniques for energy efficiency, in general, the methodologies, techniques, models, and solutions are different from this research. Even for the set of the three most correlated works, we differentiate some characteristics, as mentioned right in the sequence in which they appear.

4.5.2 Research on Dependability and Sensitivity Analysis

Smith et al. (2008) present a fault tree and Markov chain models to identify the dependency between a Blade Center System's components, presenting the failure states and the systems' availability.

- This work is the source of our research components' MTTF to represent the Blade Center. However, it differs from our research for state space-based modeling. For the Blade System representation, we use RBD without considering the dependencies between its subsystems.

Sousa (2015) presents an integrated solution composed of a methodology, methods, RBD, SPN, and CTMC models, optimization models, and a tool for planning private cloud infrastructures, considering aspects of performance, dependability, performability, and cost. The proposed optimization models are based on the GRASP metaheuristic for generating private cloud infrastructure scenarios. The model for generating performance and cost scenarios allows creating cloud infrastructure scenarios with different software and hardware configurations. The model for generating availability and cost scenarios allows creating cloud infrastructure scenarios with different redundancy mechanisms attributed to these infrastructures' components. Other works by the same author appeared as extensions of his research. Sousa et al. (2014) propose a stochastic model generator for cloud infrastructure planning that provides automatic generation of dependability and cost models for representing cloud infrastructures. A case study based on Moodle hosted on a Eucalyptus platform is adopted to demonstrate the proposed solution's feasibility (modeling strategy and tooling). Sousa et al. (2017) provide a methodology, stochastic models, and an optimization approach for assisting the planning of private cloud infrastructures, which are selected according to the availability, downtime, and cost constraints. Two case studies based on the cloud platform are adopted to demonstrate the feasibility of the proposed work.

- These works focus on solutions for private cloud infrastructures, with features more geared towards platforms and a logical redundancy mechanism, therefore, different from our purpose.

Melo (2017) proposes strategies for sensitivity analysis and the adoption of a methodology based on hierarchical modeling and study of redundancy mechanisms to assess cloud infrastructures, aiming to guarantee its quality criteria. The adopted methodology makes it possible to detect points of improvement in availability. RBDs, Markov chains, and Petri nets are used for modeling.

Júnior (2016) proposes methods for the assessment and detection of bottlenecks in cloud computing. The approach is based on hierarchical modeling and parametric sensitivity analysis techniques adapted to such a scenario. This research presents methods for building unified sensitivity rankings when different modeling formalisms are combined. In (JÚNIOR et al., 2011), the authors perform a sensitivity analysis on a computer network with redundancy mechanisms to find the system availability bottlenecks. They use Markov chains for the analytical assessment of scenarios. These two works present dependability metrics and sensitivity analysis techniques, which are common in our research. However, the approach, the assessment environment, and the models presented are different from ours.

- These last three works use the same sensitivity analysis technique; however, the networks and infrastructures represented are different from those presented in this study.

Rosendo et al. (2019) propose a methodology to automatically acquire data center hardware configuration to assess its availability through models. The proposed methodology leverages the emerging standardized Redfish API and relevant modeling frameworks. Through such approach, they analyzed the availability benefits of migrating from a conventional data center infrastructure (named Performance Optimization Data Center (POD) with redundant servers) to a next-generation virtual Performance Optimized Data center (named virtual POD (vPOD) composed of a pool of disaggregated hardware resources).

- Despite dealing with data center solutions, this work focuses on virtual solutions, different from our case, which is for physical infrastructure.

Liu et al. (2017) build up a novel-colored generalized stochastic Petri net (CGSPN) model based on IT infrastructures, which reflect the dynamic behavior and service request processing procedure under the active-active mechanism. During the modeling, the hierarchical modeling method is applied. The outline net illustrates the processing procedure, and the subnets consider the failure and resource utilization of each node.

Nguyen et al. (2019) propose a hierarchical modeling framework for reliability and availability evaluation of tree-based data center networks. The hierarchical model consists of three layers, including (i) reliability graphs in the top layer to model the system network topology, (ii) a fault-tree to model the architecture of the subsystems, and (iii) stochastic reward nets to capture the behaviors and dependency of the components in the subsystems in detail.

Nussbaumer e Liu (2013) considered specific requirements of SMEs during the selection phase of cloud service providers. The results demonstrate that security, reliability, cost, performance, flexibility, service, and support have a fundamental role and require much attention. A framework is also proposed, which focuses on a systematic service-oriented approach and helps companies analyze their existing business processes in migration.

- The focus is on performance and dependability metrics for these previous studies, but the networks and models are different from ours. Besides, EFM modeling is not presented, and neither are sustainability results like in our methodology.

4.5.3 Related Features

All the cited works have interests common to ours; however, the specific objectives are different. We created two tables to understand better the characteristics presented by the most correlated works (and main differences). Table 23 presents the description of the main characteristics analyzed in the related works. They were chosen based on what is proposed in this thesis, which does not diminish the works cited importance, as they present different purposes concerning ours, which may not have been highlighted. Table 24 shows the works that have the highest correlation to this. Note that not all works cited in the text appear in that table. This occurs for two reasons: i) for authors who have more than one work, we choose the one that is characterized as the most complete (thesis or dissertation); ii) some works cited, despite the relationship in the general context, do not have most of the characteristics presented in our approach.

Table 23 – Description of the Features Analyzed.

Feature	Description
F1	Electrical architecture
F2	IT architecture
F3	Cooling architecture
F4	Systematic Literature Review
F5	Energy demand proposal
F6	EFM
F7	RBD
F8	SPN
F9	CTMC
F10	FT or GSPN or CGSPN or other
F11	Sustainability analysis
F12	Dependability analysis
F13	Sensitivity analysis
F14	Maintenance analysis

Source – Elaborated by the author.

4.6 FINAL CONSIDERATIONS

This chapter presented the process used for the SLR, identifying the relevant works for the research context. The applied search strings returned 3,295 papers in three search bases for the five research questions. We presented research questions, search strings, quantitative and qualitative results, and the primary evidence for topics of interest, and the problems and opportunities for new research were identified. After the works' data selection and extraction phases, in which we selected a total of 57 studies for this SLR, the extracted and summarized data presented evidence about the contexts covered. However, the results proved that the essence

Table 24 – Analyzed Features.

Work	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
This Thesis	■	■		■	■	■	■	■			■	■	■	■
Júnior (2019)	■					■	■	■	■		■	■		
Callou (2013)	■		■			■	■	■	■		■	■		
Sousa (2015)		■					■	■	■			■		■
Souza (2013)		■	■		■		■	■				■		■
Melo (2017)		■					■	■	■			■	■	
Figueirêdo (2011)	■					■	■				■	■		
Júnior (2016)		■					■	■	■			■	■	
Marwah et al. (2010)	■							■			■			
Banerjee et al. (2012)	■									■	■			
Sankar, Gauthier e Gurumurthi (2014)	■									■	■	■		
Liu et al. (2017)		■						■		■		■		
Silvaa et al. (2014)	■						■	■				■		
Müller e Strunz (2012)	■		■							■		■		
Nguyen et al. (2019)		■								■		■		
Ritchie e Brouwer (2018)	■						■					■		
Rosendo et al. (2019)		■					■					■	■	
Bennaceur e Kloul (2020)	■	■	■							■		■		

Source – Elaborated by the author.

of this research's objective has remained original and unprecedented. Additionally, some works were presented that have the same context as our research. Two main themes were presented to frame the studies. In most cases, works that are interested in sustainability do not focus on dependability and vice versa. In the end, two tables were shown to summarize the most correlated works' features.

5 MODELING

This chapter presents a set of models proposed to plan electrical and network architectures of the data centers according to the physical components' redundancy classification, given by tier. Our proposal is based on a hierarchical and heterogeneous approach focusing on aspects of sustainability and dependability. Initially, a proposal is presented to calculate the electrical demand for data center projects, and then the models are presented. Although a data center project consists of several subsystems, we represent only electrical and network architectures on models.

5.1 MODELS FOR ELECTRICAL ARCHITECTURES

The modeling represents the use of mathematical techniques to describe a system's functioning or part of a productive system that can be evaluated via simulation. There is no need to have an implanted system, so it is possible that in the design phase, the main points of faults are identified and receive special attention, aiming to increase the quality-of-service levels for the architecture.

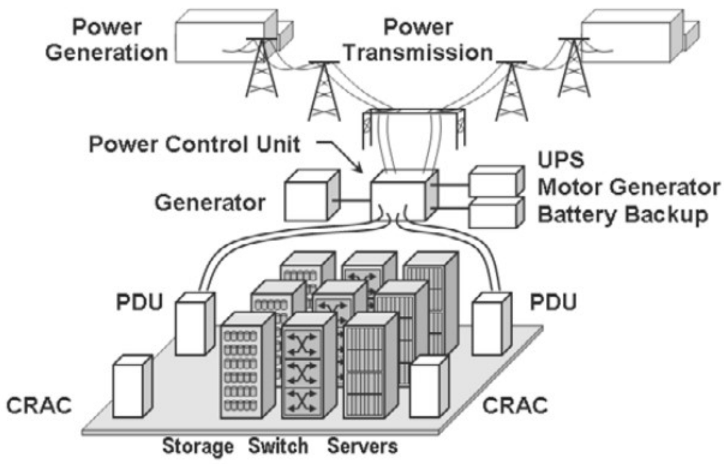
Given the demand for new data center infrastructures, one of the critical issues for projects and implementation is focused on spaces, especially since there is a need to dimension the various subsystems involved. These can be classified as electrical systems, support spaces, and a computer room (MARIN, 2011).

The reliability and serviceability of electrical and cooling systems are fundamental design requirements to enable IT and cooling systems' successful operation. The equipment and subsystems depend entirely on electrical distribution, starting with the transformers of the concessionaire and the building, switches, uninterruptible power supplies (UPS), power distribution units (PDUs), and power supplies, feeding the fans and the internal components of the IT equipment. These components will have a degree of inefficiency, resulting in converting electricity to heat (loss of energy) (GENG, 2014).

Even during a power outage, however short, it is crucial to maintain the data center's availability, so before generators are started and ready to provide stable power, UPS are used. Power management switches allow automatic transfer from the primary utility to the backup source. PDUs transform the high voltage power provided by power management systems into IT equipment where needed. Electric power cables can be found under raised floors or in suspended transport systems. For example, PDUs can be located in a data center in different zones, providing power to servers, storage, and network devices along with heating, ventilating, and air conditioning (HVAC) and computer room air conditioning (CRAC) units (SCHULZ, 2016).

Figure 16 shows the electrical arrangement and its correlation with the other subsystems that make up a data center. The UPS supplies short-term power using batteries or a motor-generator until the backup generators are ready for use. A motor-generator draws energy from the utility in a primary power loss, providing enough kinetic energy to run a small generator long enough for gas, diesel, propane, or natural gas generators (SCHULZ, 2016).

Figure 16 – Powering a Data Center or Habitat for Technology.



Source – (SCHULZ, 2016, p. 134)

To support the load initially installed and with some slack for future growth, it is essential to design a data center's electrical distribution correctly. There is no standard to be followed; but mainly due to issues related to sustainability, increased energy costs, greenhouse gas emissions, and density of energy use by chassis racks, it is vital to know the data center subsystems and their consumption so that it is possible accurately dimension the total load. In this work, we consider the IT critical load, cooling and battery system, lighting, and space for future growth (current support space) to perform an electrical load estimate for four different data center architectures, according to the specifications of redundancy for the classification of tiers 1, 2, 3 and 4.

The following subsection presents the equations for calculating data center subsystem loads, the results of which serve as input parameters for EFM models.

5.1.1 Proposed Energy Demand

To evaluate the data center's energy demand, we performed some calculations based on recommendations for more efficient projects presented from Equation 5.1 to 5.6. For example, we consider the distribution of load per rack, rather than a distribution per square meter, for estimating IT load. It is worth noting that all equations are based on recommendations given by Marin (2011). Thus, the values of the constants are justified based on experiences of energy-efficient projects.

According to Marin (2011), the average chassis consumption was calculated from different manufacturers, from which we take the four highest values in kW to make the chassis racks' average consumption that represents the IT load. In this direction, we obtained an average consumption of 14.125 kWh per chassis rack. We consider that each rack has five chassis (based on a particular manufacturer). Table 25 shows the number of chassis, chassis racks, and the total area for each tier of the data centers represented for this study. Equation 5.1 shows how to calculate the IT load.

Table 25 – Amounts of Chassis, Racks and DC Area by Tier.

Quantities/Area	**Tier 1**	**Tier 2**	**Tier 3**	**Tier 4**
Chassis	75	150	300	600
Racks	15	30	60	120
Total area (m^2)	400	600	800	1000

Source – Elaborated by the author.

$$Rack_{load} = (\frac{Amount_{Chassis}}{5}) \times RackAverage_{consumption} \tag{5.1}$$

We also applied a growth factor as a percentage of 10% on the IT load over the five years (MARIN, 2011). This is to be represented as the margin available for new equipment added in the future without stress for the electrical load. For our study, we consider that this load currently represents the support space (support rooms), which can be understood as the spare equipment's storage rooms. To calculate the ambient lighting, we consider that the support spaces' consumption will be 1.5% of the IT load consumption since this space is little used. Furthermore, the lighting of the IT room is done considering the total area, in square meters. So, we have the Equation 5.2.

$$Full_{lighting} = (Rack_{load} \times 0.015) + (\frac{Area_{IT}}{100}) \tag{5.2}$$

We also consider the consumption of battery charges to be 15% of the IT load (MARIN, 2011). Furthermore, the inefficiency factor as being 10% of the IT load (MARIN, 2011). So, we have the Equation 5.3.

$$Batteries_{load} = (Rack_{load} \times 0.15) + (Rack_{load} \times 0.10) \tag{5.3}$$

To determine the cooling system's load, we need to know DC's partial load, considering the subsystem evaluations presented so far. So, we have the Equation 5.4.

$$DC_{load} = Rack_{load} + Full_{lighting} + Batteries_{load} \tag{5.4}$$

The load consumed for cooling the data center, including pumps, compressors, and chillers, represents one of the most massive energy consumptions in the data center. The efficiency of which is around 65% (Cooling_Factor) (MARIN, 2011). This value, therefore being considered as the cooling factor. So, we have the Equation 5.5.

$$Cooling_{load} = DC_{load} \times Cooling\_Factor_{load} \tag{5.5}$$

Finally, we estimate the total load for the data centers represented in this study with the Equation 5.6.

$$Full\_DC_{load} = DC_{load} + Cooling_{load} \tag{5.6}$$

Table 26 shows the loads achieved for the subsystems represented for each DC tier. The load completed for each DC tier will serve as an input parameter for the energy flow models (EFM) for each corresponding level. We will determine the total electrical demand based on the efficiency of the components used from these values.

Table 26 – Subsystem Loads by DC Tier.

Subsystems	**Tier 1 (kWh)**	**Tier 2 (kWh)**	**Tier 3 (kWh)**	**Tier 4 (kWh)**
IT (Chassis Rack)	211.86	423.75	847.50	1695.00
Space to grow (Support rooms)	129.35	258.70	517.41	1034.81
Full illumination	7.18	15.36	28.71	50.43
Batteries	52.97	105.94	211.88	423.75
Cooling	260.90	552.43	1043.57	2082.60
Full DC	**662.26**	**1326.18**	**2649.07**	**5286.59**

Source – Elaborated by the author.

5.1.2 Input Parameters

A data center's electrical distribution system must guarantee a continuous and safe option since the entire critical IT load depends on it. The more sophisticated and redundant these systems are, the greater the load to feed them. In the energy flow model, integrated into the Mercury Tool (MACIEL et al., 2017), several components are offered for the user to model.

The following are the components that must be part of a power architecture (MARIN, 2011) of our work.

- ACSource - Represents the energy utility.
- Generator - Alternative energy source, in case of lack of the concessionaire.
- ATS (Automatic Transfer switch) - It is typically used to switch distribution timing to the generator in the event of a local utility lack.
- Subpanel - Electrical circuit breaker box that serves as a point for energy distribution.
- UPS (Uninterruptable Power Supply) - It is a set of batteries that supply power to the data center in the event of a power failure by the source.
- SDT - The transformer is responsible for stabilizing the energy that arrives to leave it at the correct power and current for the equipment.
- Junction Box - Are boxes where electrical cables are distributed to the racks.
- Power Strip - Allows multiple components to be powered by a block of electrical outlets, which connect to the end of a flexible cable.

The computed electric demand for each DC tier will serve as an input parameter for the energy flow models (EFM). From these values, we will receive the total electrical demand, based on the efficiency of the components used, representing the minimum amount of energy that must be received from the assignee to power the data center subsystems.

From the graphical interface of the Mercury (MACIEL et al., 2017), the user can choose among the various components available for EFM to those that best represent the desired architecture. Besides, for each component metric, we can use the default values or change them as needed. For this study, we changed the maximum energy capacity and the price of all components, leaving the default values of efficiency and embedded energy (Em. En.). Component prices have been updated to the most current reality. The parameters of efficiency, embedded energy, and price are the same for the different tiers, differently from the maximum energy with different values. Each component's maximum energy capacity was calculated based on the electrical demand (See Table 26). The input parameters for each component are shown in Table 27.

To perform the dependability analysis of electrical architectures, we had to implement RBD modeling. We try to be faithful to EFM, but some possible differences between the components' connections can be noticed due to each modeling technique's nature. Unlike EFM,

Table 27 – EFM Model Parameters.

Components	Effi. (%)	Em. En. (GJ)	Price (US$)	Max. Power (kW)			
				Tier 1	Tier 2	Tier 3	Tier 4
ACSource	95.30	111.000	8250.00	3774.92	7559.24	15099.68	30133.52
Generator	25.00	97.700	36300.00	50332.26	100789.92	201229.04	401780.28
$\text{Generator}_{redun.}$	25.00	9.700	3300.00	–	–	201229.04	401780.28
ATS	99.50	0.799	440.00	993.40	1989.27	3973.60	7929.87
Subpanel	99.90	0.428	110.00	728.49	1458.80	2913.97	5815.24
UPS	95.30	61.040	33000.00	3774.92	7559.24	15099.68	30133.52
$\text{UPS}_{redundant}$	95.30	3.139	1980.00	–	7559.24	15099.68	30133.52
SDTransformer	98.50	0.359	302.50	1655.67	3315.46	6622.67	13216.46
JunctionBox	99.90	0.267	30.00	728.49	1458.80	2913.97	5815.24
PowerStrip	99.50	0.356	40.00	993.40	1989.27	3973.60	7929.87

Source – Elaborated by the author.

for the representation of RBDs models in Mercury Tool, only two input parameters are required for each block, MTTF and MTTR, beyond defining the type of block for better representation of the system modeled. For all components, in all tiers, we consider the 8-hour MTTR (CALLOU, 2013). For the sensitivity analysis, we varied this value from -50% to +50%, that is, between 4 and 12 hours. MTTFs were obtained from (CALLOU et al., 2014; FERREIRA et al., 2020). Only the junction box component suffered a reduction of 0.8, considering the MTTF value. Due to the manufacturer's original value, this reduction can be applied at a 95% confidence level, as Smith (2017). Table 28 shows the MTTFs of the components and the variation of the value to -50% to +50%, to be used in the sensitivity analysis.

Table 28 – Input Parameters of RBD Models.

Components	MTTF (h)	MTTF variation	
		-50%	+50%
ACSource	4,380	2,190	6,570
Generator	2,500	1,250	3,750
ATS	24,038	12,019	36,057
Subpanel	152,000	76,000	228,000
UPS	50,000	25,000	75,000
SDT	141,290	70,645	211,935
JunctionBox (JB)	522,400	261,200	783,600
PowerStrip (PS)	215,111	107,556	322,667

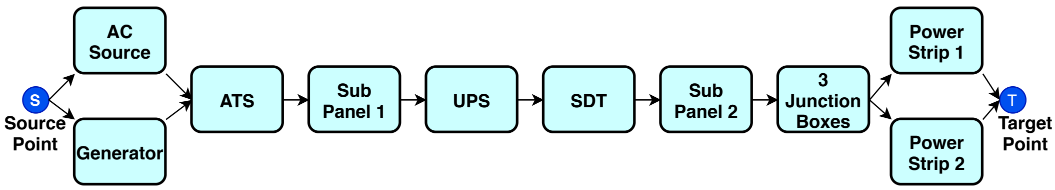
Source – Elaborated by the author.

5.1.3 Proposed Models

Initially, we modeled four architectures to represent the energy flow models for data centers tiers 1, 2, 3, and 4. The $TargetPoint$ and $SourcePoint$ (See Figure 17) components represent the IT power demand and the power supply demand, respectively. According to the tier classification, $TargetPoint$ receives the value representing the total load of the data center project, whose values are presented in Table 26 (last line). The $SourcePoint$ component does not receive an input value since this value is estimated after evaluating the model.

Figure 17 presents the electrical architecture of a tier 1 data center. This model has the minimum necessary to be considered a DC electrical system. Under normal operating conditions, the electrical system is supplied by the local utility. The energy flow leaves the source/concessionaire ($ACSource$), passing through the low voltage panel ($Subpanel1$), through the uninterruptible power supply (UPS), through the power distribution units – composed of a transformer (SDT) and a subpanel, through junction boxes, power strips and arrives at the power distribution units for the IT devices (Chassis Racks). For this architecture, we consider three junction boxes, each connected to two power strips. In the absence of power supply by the concessionaire, the generator takes over the load, and, during the switching period (when ATS starts the generator), the UPS system keeps the IT critical load in operation.

Figure 17 – EFM for Tier 1 Data Center (EFM T1DC).



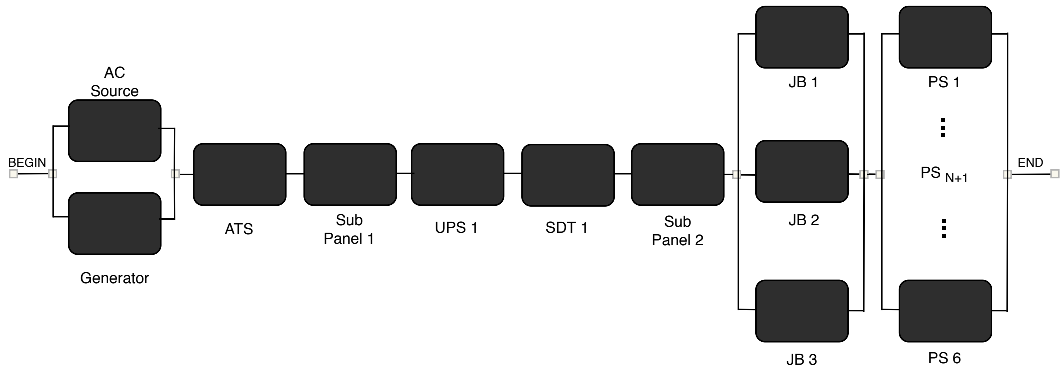
Source – Elaborated by the author.

For the RBD modeling of electrical architecture, we created an array of parallel-series blocks. If there is no parallel block in case of failure of serial blocks, there will be a break in electrical services' continuity. In this sense, for tier 1 models, we are more likely to stop the site where there is almost no component redundancy. Figure 18 presents the RBD model for this electrical architecture. Note that for EFM, for simplicity, we represent each junction box connected to two power strips, while in the RBD model, we represent this by parallel blocks.

For modeling the electrical architecture of tier 2, we consider meeting all the specifications of tier 1 plus the redundancy of the components that make it tier 2, e.g., UPS and PDU. Figure 19 shows the electrical model for a tier 2 data center, characterized by the presence of redundant components. In this model, in addition to the central UPS ($UPS1$), there is a redundant module ($UPS2$), which assumes the IT critical load in case of failure of the main one. In this case, the electric flow will go through the redundant PDU ($SDT2$ and $Subpanel2$). We consider four junction boxes, each connected to eight power strips, to support the IT load in the face of UPS1 failure. For the other components, the same explanation of the tier 1 model is valid.

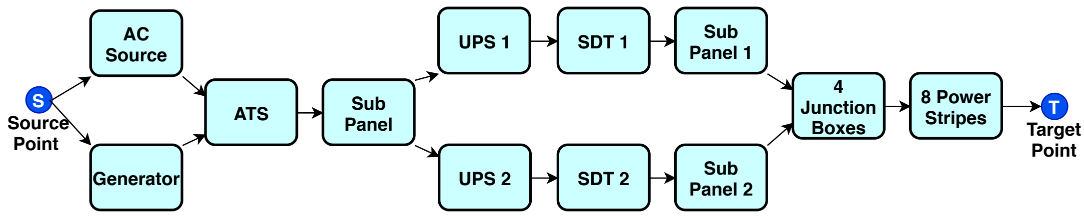
For the RBD modeling of tier 2 electrical architecture, we created an array of parallel-series blocks, in which the parallel blocks represent a higher level of redundancy than tier 1. In the same way as in tier 1, in the case of serial block failure in which there is no parallel block,

Figure 18 – RBD for EFM Tier 1 Data Center (RBD-EFM T1DC).



Source – Elaborated by the author.

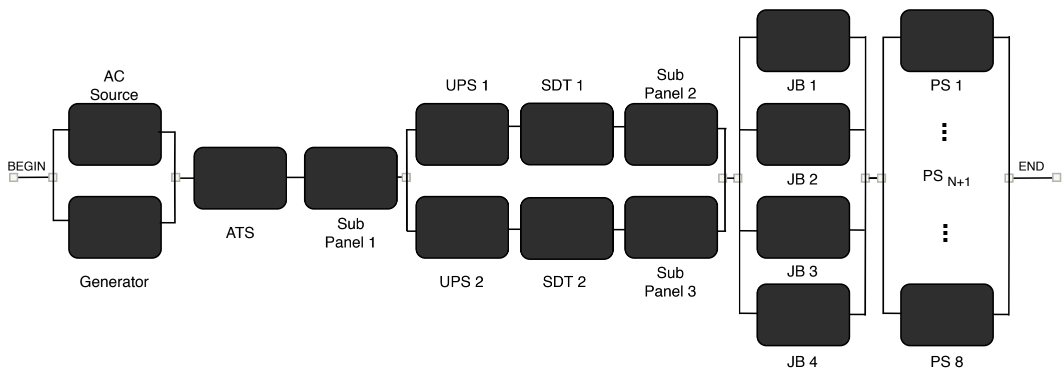
Figure 19 – EFM for Tier 2 Data Center (EFM T2DC).



Source – Elaborated by the author.

there will be a break in electrical services' continuity. However, only the ACSource, Generator, ATS, and Subpanel 1 components do not have redundant components for this tier. However, in terms of service provision, the generator feeds the site in the absence of ACSource. Figure 20 presents the RBD model for this architecture.

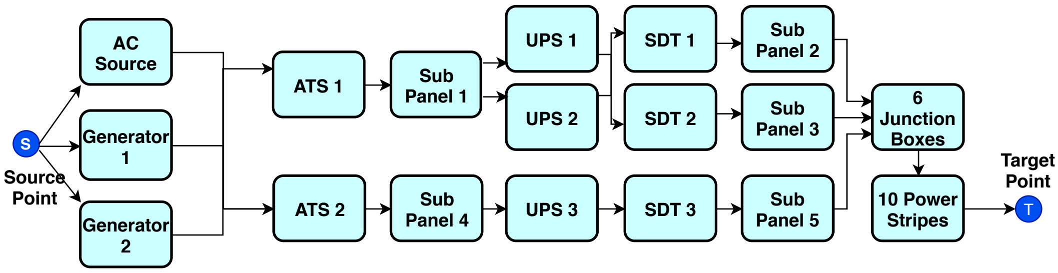
Figure 20 – RBD for EFM Tier 2 Data Center (RBD-EFM T2DC).



Source – Elaborated by the author.

To model the electrical architecture of tier 3, we consider meeting all the specifications of tier 2. For specific components more critical for the site's continuous supply, according to redundancy specifications, we added another simple distribution path concerning the path main. Figure 21 shows the electrical architecture for a tier 3 data center, characterized by the presence of an alternative distribution path. So, maintenance services are carried out without interrupting the operation. In this model, each path has automatic transfer switches (ATS) capable of switching the critical load to any distribution paths and any sources and generators. In addition to the main generator ($Generator1$), there is a redundant module ($Generator2$). The main path ($ATS1$) has a redundant module for UPS ($UPS2$), transformer ($SDT2$), and subpanel ($Subpanel2$). In normal operation, the critical load is fed through the main path ($ATS1$) by the local utility ($ACsource$). However, in the absence of $ACsource$ supply, the synchronized switches $ATS1$ and $ATS2$ switch the main path to the generators to feed the IT critical load. We consider six junction boxes for this architecture, each connected to ten-line filters, to support the IT load.

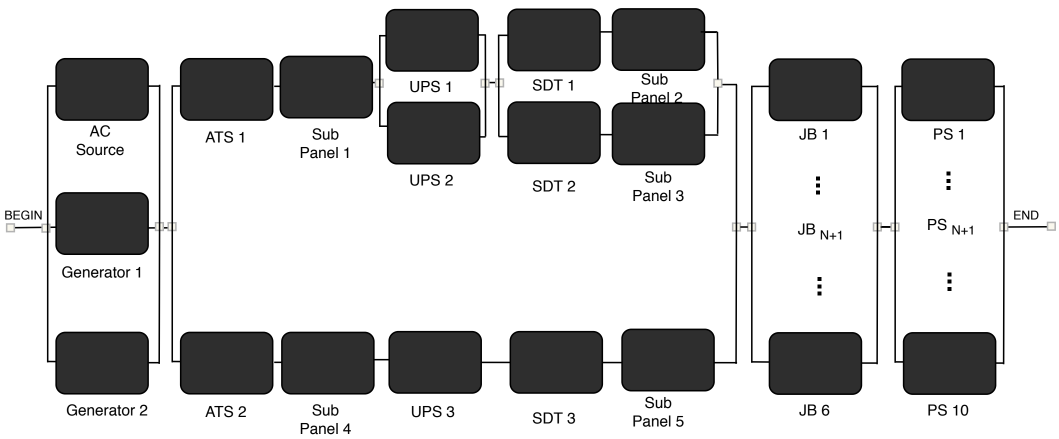
Figure 21 – EFM for Tier 3 Data Center (EFM T3DC).



Source – Elaborated by the author.

For the RBD modeling of tier 3 electrical architecture, we created an array of parallel-series blocks, in which for the failure of a block, there will be at least one other substitute. Note the presence of a redundant generator it is more difficult for the system to become inoperative, given the presence of a secondary distribution path, which can be used in case of failures in the main path or scheduled maintenance actions. Figure 22 presents the RBD model for EFM Tier 3.

Figure 22 – RBD for EFM Tier 3 Data Center (RBD-EFM T3DC).

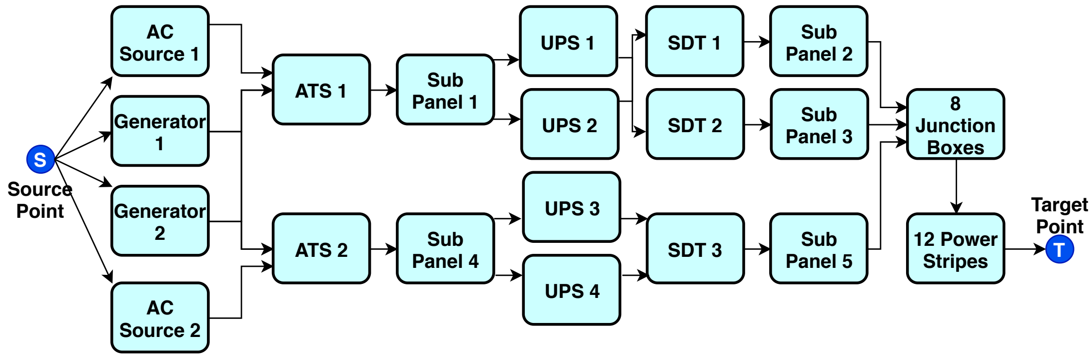


Source – Elaborated by the author.

The electrical system of tier 4 is the most robust of all. In addition to meeting the needs of tier 3, several other options for redundancy of components and distribution paths exist. This makes them more fault tolerant as well as providing greater availability. In addition to the characteristics of tiers 1, 2, and 3, this tier includes the characteristics of redundant energy (primary and secondary) and cooling, careful site selection, alternative locations for business continuity or disaster recovery. The physical facility is protected from the risks of applicable threats, including tornadoes, hurricanes, floods, or other acts of nature.

Figure 23 shows the electrical architecture for a tier 4 data center, characterized by the ability to recover from faults automatically. The main difference between this power diagram and that of a tier 3 data center is that two electrical distribution paths are simultaneously active, which are provided from different sources. All components are powered, and IT critical loads have dual sources (at least), each connected to a different distribution path. In the absence of one or both sources, the generators are activated, and the loads remain in operation by either (or both) of the distribution branches.

Figure 23 – EFM for Tier 4 Data Center (EFM T4DC).



Source – Elaborated by the author.

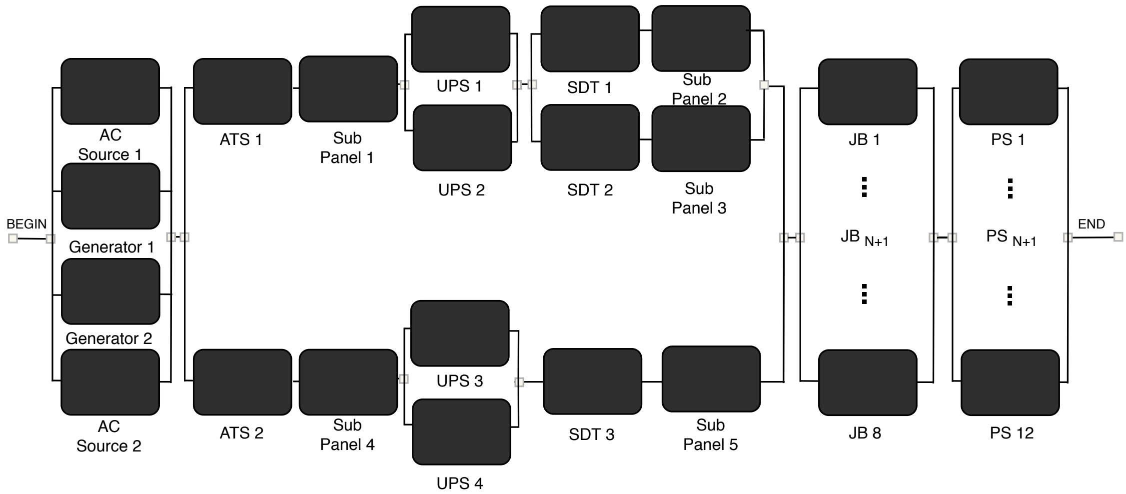
For the RBD modeling of tier 4 electrical architecture, we created an array of parallel-series blocks, in which for the failure of a block, there will be at least one other substitute. In addition to a redundant generator, there is also redundancy for ACSource, representing a minimum of two primary energy sources. In the secondary distribution path (ATS2), there is a redundant module for UPS3 (UPS4) when compared to tier 3. Figure 24 presents the RBD model for tier 4 architecture.

We choose the minimum recommended for each tier for all electrical architecture models, but the components, arrangements, and quantity can vary according to the project's needs in the real world.

5.2 MODELS FOR IT ARCHITECTURES (NETWORKS)

Data center networks can contain multiple components, several protocols, and different technologies, which can complicate configuration details. In this sense, the hierarchical layered model can help mitigate this complexity by facilitating implementation and maintenance. We consider the three layers model for the modeling of data center network architectures as presented

Figure 24 – RBD for EFM Tier 4 Data Center (RBD-EFM T4DC).



Source – Elaborated by the author.

in Section 2.1.3. All the tiers DC aggregate the three layers presented. In the access layer are the chassis racks that represent the network storage endpoints. The aggregation layer links the access and core layers. The core layer represents the high-speed connection point between the Internet Service Provider (ISP) and external networks for incoming and outgoing data flow. We decided to use the blade system chassis due to its energy and space-saving characteristics, promoting faster maintenance. Such characteristics corroborate the current interests in data center design. Because the chassis already have integrated switches, we abstract from the modeling the use of layer two switches (L2 switches).

It is important to note that the components used to represent the models in this work and the models themselves are generic enough to be adapted for different needs. Furthermore, they are independent of the manufacturer's brand. We seek to represent models and scenarios according to the definition of use in real data centers (IV et al., 2006; VERAS, 2009). Our focus is on representing internal networks and in the physical structure.

To modeling the data center networks, we consider some assumptions for good high availability network designs, an example of which is the number of particular components with a margin for growth. We assume that the networks have more robust and powerful equipment, so it is possible to reduce the amount and achieve a greater density of spaces. For example, suppose the use of 48-port routers and switches, which is sufficient to meet the current needs of networks and still support expanding the access layer without adding equipment in the core and aggregation layers. As downtime tends to be very expensive, we try to project the levels of physical redundancy for each tier, respecting each classification's minimum specifications.

We discard the information from the logical connections and the "daisy chain" between the switches and do not specify the number of uplink ports. We defined communication protocols, and link redundancy protocols are not part of the scope of our modeling.

We consider having stub networks for tiers 1, 2, and 3, with a single connection to the ISP and a single outbound path. A failure in the ISP will cause the networks to fail. Due to the need for high availability, we consider that there must have a use percentage of 60%, 70%, 80%, and 90% for tiers 1, 2, 3, and 4 for all K-out-of-N (KooN) blocks, respectively. These percentages were determined to maximize the availability requirements according to the tier classification.

To represent the data center networks, firstly, we model a blade center system (see Figure 6). Initially, we present a model to the blade server components, and then we model a chassis, which is composed of fourteen blade servers. This system was chosen to represent the networks' access layer as they offer the greater density and efficiency (Smith et al., 2008), which are highly desired requirements for current and future networks. It is essential to note that we do not consider dependencies between subsystems, so we only perform modeling at a high level of abstraction.

As a characteristic of our hierarchical approach, the MTTF resulting from the first model (RBD-Net - Server, Figure 25) is used in the second model (RBD-Net - Chassis, Figure 26) to represent server failure. The MTTF resulting from the second model, which represents a Chassis blade, is used in the third model (RBD-Net - Rack, Figure 27) to represent the number of chassis per rack. With this third model's evaluation, we obtained the MTTF to represent the racks in the other networks of tiers 1, 2, 3, and 4 (Figures 28 to 31). Finally, the MTTF resulting from each of these networks is used in the SPN model (Figure 32) to verify compliance with levels of services presented in different SLA categories, given the proposed maintenance policy's characteristics.

5.2.1 Input Parameters

Following, we present the input parameters corresponding to the RBD models of the data centers' IT architectures.

Table 29 presents the input parameters of the "Blade Server RBD Model". For each component, the MTTF value source is presented, the block acronym used in the model, the component's name, the MTTF (in hours), and the variation MTTF to be used sensitivity analysis (-50% to +50%). Note that the components that have a * have experienced a 0.8 reduction in the MTTF value. This reduction can be applied to a confidence level of 95 % (SMITH, 2017) as already mentioned for the junction box component in Subsection 5.1.2. Some components obtained from the Smith et al. (2008) suffered this reduction because they had an unusually high MTTF, and the cited paper was made in partnership with a manufacturer.

Table 30 presents the components of "Chassis Blade RBD Model". For this model, we created seven blocks in series, five of which are of the KooN type. The same description given

Table 29 – Input Parameters for Blade Server RBD Model.

Source	Ab.	Component Name	MTTF (h)	MTTF (h) -50%	MTTF (h) +50%
(Smith et al., 2008)	BB	Base Blade	220,000	110,000	330,000
	CP	CPU*	50,000	25,000	75,000
	DM	Memory Bank (DIMMs)	480,000	240,000	720,000
	HD	Hard Disk Drive	200,000	100,000	300,000
	FC	Fiber Daughter Card*	260,000	130,000	390,000
	EC	Ethernet Daughter Card*	1,240,000	620,000	1,860,000

Source – Elaborated by the author.

for the data in Table 29 is valid for this one. For the Blade Server component, shown in the last line (Table 30), the MTTF source is the first RBD model, i.e., Blade Server.

Table 30 – Input Parameters for Chassis Blade RBD Model.

Source	Ab.	Component Name (KooN)	MTTF (h)	MTTF (h) -50%	MTTF (h) +50%
(Smith et al., 2008)	MP	Midplane	310,000	155,000	465,000
	SW	Software	17,520	8,760	26,280
	PS	Power Supply (2/4)	670,000	335,000	1,005,000
	FS	Fiber Channel Switch (1/2)	320,000	160,000	480,000
	ES	Ethernet Switch (1/2)	120,000	60,000	180,000
	BW	Blower* (1/2)	620,000	310,000	930,000
RBD-Net - Server	BS	Blade Server (6/14)	27,562	13,781	41,343

Source – Elaborated by the author.

All components belonging to the first and second models received an MTTR of 4-hours (CAMBOIM et al., 2020b). We consider this time due to the ease of repair or replacement of components, should they fail, which is directly associated with the maintenance policies that should be adopted. For the sensitivity analysis, its variation of -50% and + 50%, i.e., 2 and 6, respectively.

Table 31 shows the component parameters for the other models. For the third model, which aims to represent the mean time to failure of a rack composed of five chassis, we use the MTTF calculated from the second model (RBD-NET - Chassis). We use the MTTF calculated from the third model's evaluation (RBD-NET Rack) to represent the four data center networks' models. QSFP, SFP, and UTP are the types of transceivers used to connect network components. For Blade Chassis and transceivers, the MTTR is 4-hours (CAMBOIM et al., 2020b) due to ease of replacement. For chassis racks, routers and switches, the MTTR is 8-hours (CAMBOIM et al., 2020b). Note the components with *, which were the ones that suffered the MTTF reduction (as explained previously).

Table 31 – Input Parameters for RBDs Network Models.

Source	Ab.	Component Name	MTTR (h)	MTTF (h)	MTTF (h) -50%	MTTF (h) +50%
RBD-Net - Chassis	CB	Chassis Blade	4	12,634	6,317	18,952
(EXTREME, 2020)	QSFP	Quad Small Form Factor Pluggable*		1,014,000	507,000	1,521,000
	SFP	Small Form Factor Pluggable*		980,200	490,100	1,470,300
	UTP	Unshielded Twisted Pair*		240,000	120,000	360,000
RBD-Net - Rack	CR	Chassis Racks	8	28,849	14,424	43,273
(CISCO, 2020a)	RT	Router*		96,154	48,077	144,231
(CISCO, 2020b)	SW	Switch (Layer 3)*		88,684	44,342	133,027

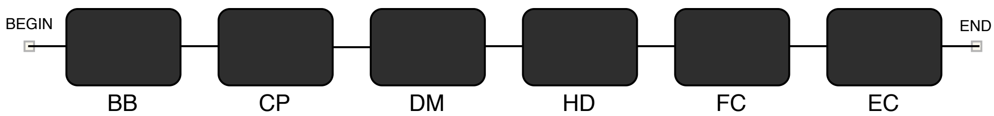
Source – Elaborated by the author.

5.2.2 Proposed Models

The first three models were created to represent the access layer components of data center networks. Initially, we describe the blade center system (first two models), and then we represent a rack composed of five chassis.

The first model represents the components of a blade server (to compose a chassis). This model has the components connected in series because a component's failure represents the system's failure (server blade). Figure 25 shows the first RBD model, whose acronyms and MTTFs of the components were presented in Table 29.

Figure 25 – RBD for Networks - Blade Server (RBD-Net - Server).



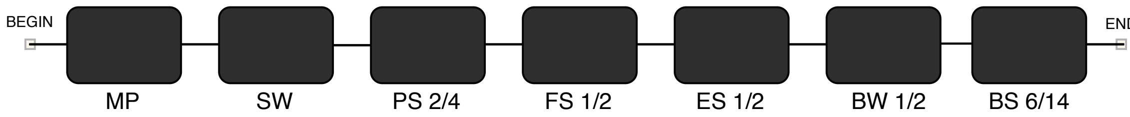
Source – Elaborated by the author.

The reliability of this system is calculated using Equation 2.16.

It is worth mentioning that we followed the description given in Smith et al. (2008) to compose the two subsystems representing a blade system. However, it was not our interest to be based on the dependence and failure characteristics between the components and subsystems (as is modeled in the mentioned paper). Therefore, we perform this model to reach the MTTF of the blade center system and compose our scenarios.

The second RBD model represents a chassis (RBD-Net Chassis), which has midplane (MD), software (SW), four power supply (PS 2/4 - at least two must work), two switches with fiber optic link (FS 1/2 - at least one must work), two switches with an Ethernet link (ES 1/2 - at least one should work), two blowers (BW 1/2 - at least one must work) and fourteen servers (BS 6/14 - at least six must work). The choice of six servers in operation is due to six being connected to two primary power supply and eight being connected to the other two secondary power supply. Figure 26 shows the second RBD model, whose MTTFs were presented in Table 30. The model's general structure has represented in series with some non-identical KooN blocks.

Figure 26 – RBD for Networks - Chassis Blade (RBD-Net - Chassis).

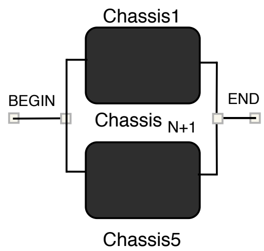


Source – Elaborated by the author.

Equations 2.16 and 2.19 calculates the reliability of this system.

Figure 27 presents the third RBD model (RBD-Net - Rack), representing that each rack comprises a five-chassis. This way, we obtained the MTTF of the racks that will compose the DC networks (from Tier 1 to 4). The structure comprises five chassis parallel blocks (reduced figure), with each block has the same MTTF and MTTR values. The values used for MTTF is 12,634-hours (seen Table 31) and MTTR is 4-hours (CAMBOIM et al., 2020b). The reliability of this system is calculated using Equation 2.18.

Figure 27 – RBD for Networks - Chassis Rack (RBD-Net - Rack).

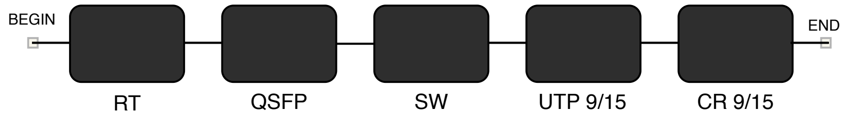


Source – Elaborated by the author.

Subsequently, we present the RBD models of the IT architectures of tier 1 to 4 data centers, represented by the Figures 28, 29, 30, and 31. In all the following networks, the MTTF value achieved with the third RBD model (Figure 27) represents the racks that make up the access layer. To calculate the reliability of the networks, different arrangements of Equations 2.16, 2.19 and 2.18 are made.

Figure 28 presents the RBD model for the tier 1 DC network (RBD-Net - T1DC). The system is represented by series blocks of the simple type and the last two blocks of the KooN type. In the core layer, we have a router that connects to the aggregation layer via QSFPs transceivers. We have a layer three (L3) switch for the aggregation layer, which connects to the fifteen chassis racks via UTPs transceivers. The downstream flow happens from the core layer to the access layer, and the opposite occurs with the upstream flow (which is the same for the four networks). This tier 1, which has fifteen chassis racks, a minimum of nine (60%), must be running simultaneously for the system to be in functional mode.

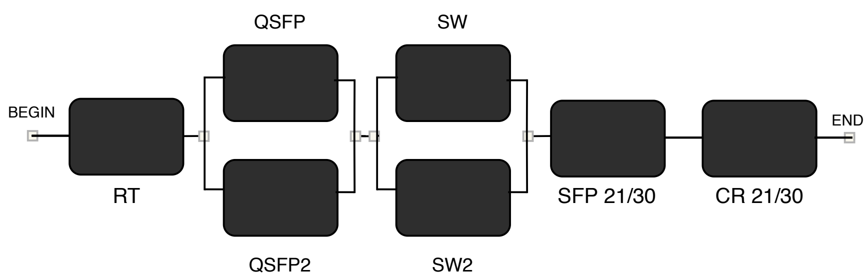
Figure 28 – RBD for Networks - Tier 1 Data Center (RBD-Net - T1DC).



Source – Elaborated by the author

Figure 29 shows the RBD model for the tier 2 DC network (RBD-Net - T2DC). The difference between this and the RBD-Net - T1DC is we have doubled the number of components, except for the router. The system is represented in series by simple parallel blocks and KooN. In the central layer, we have a router that connects to the aggregation layer through QSFPs transceivers connected to two L3 switches. We have 30 chassis racks in the access layer interconnected to the aggregation layer via SFPs optical fiber links. For tier 2, at least 21 of 30 (70%) chassis racks must be running simultaneously.

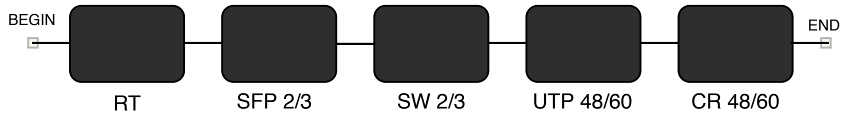
Figure 29 – RBD for Networks - Tier 2 Data Center (RBD-Net - T2DC).



Source – Elaborated by the author.

Figure 30 shows the RBD model for the tier 3 DC network (RBD-Net - T3DC). Blocks represent the system in series, one of the same type, and the other blocks are the KooN type. We have a router connected to three aggregation switches in the central layer, which interconnects 60 racks via UTPs transceivers. Tier 3 network must have 80% of chassis racks running simultaneously to be considered in operation.

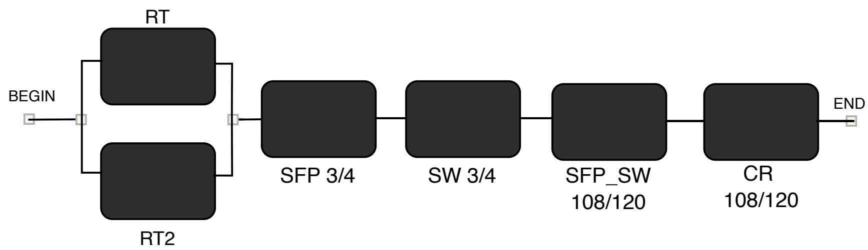
Figure 30 – RBD for Networks - Tier 3 Data Center (RBD-Net - T3DC).



Source – Elaborated by the author.

Figure 31 shows the RBD for the tier 4 DC network (RBD-Net - T4DC). Specifically, for this tier, we added router redundancy and increased the physical links' capacity. Thus, we have a multi-homing stub network with two ISPs. Therefore, this network has a more remarkable ability to tolerate failures in their providers' logical/physical connection. The system is represented by a series of parallel and KooN blocks. Within these possibilities, we can have a primary link (active) with one or more links in a state of passive redundancy (e.g., hot-standby), or we can have all links active implementing some form of load balancing. However, the logical dependencies of active and passive states cannot be represented by RBDs. We have four L3 switches connected to routers via SFPs transceivers and interconnect the 120 access layer racks via SFPs. Tier 4 network must have a minimum of 90% chassis racks running simultaneously. Otherwise, the system enters a failed state.

Figure 31 – RBD for Networks - Tier 4 Data Center (RBD-Net - T4DC).



Source – Elaborated by the author.

5.3 SPN MODEL FOR MAINTENANCE POLICY

The SPN model of this thesis was designed to represent the policy of maintenance proposed to assess the availability of the four tiers data centers following procedures adopted to minimize the occurrence of infrastructure failure events or recovery after a stoppage.

The values used for the SLA categories, corresponding to the intervals between maintenance and repair times, are based on real needs given the demands for today's data centers' high availability.

In addition to the redundancy capacity being responsible for increasing the system's availability, making it fault-tolerant, maintenance actions are of equal importance to maintain the operating system's state.

The maintenance policy approaches correspond to preventive maintenance and corrective maintenance performed to increase the operational time or restore a system from a faulted state to a functioning state.

We consider that the failed items will be recovered or replaced, and there will be an administrative time associated with the team call. The behavior of the system after the repair action will depend on the type of maintenance performed.

In the periodic preventive maintenance policy, maintenance is performed periodically at fixed time intervals kT (k = 1, 2, ...), regardless of the history of component failures (EBELING, 1997). For this work, preventive actions are planned in three different categories that correspond to different intervals between their realization, being "little managed," whose realization happens every 720 hours, "managed" has actions every 360 hours, and "well managed" with maintenance every 168 hours.

When an item fails, it enters the repair process. The repair process itself can be broken down into several different subtasks and delay times. Supply delay is the total delay time in obtaining necessary parts or components to complete the repair process. This time can consist of administrative, production, or delivery deadlines for acquisition, repair of subcomponent failures, and transportation times. In general, this time is influenced by the complexity of selecting spare parts and components available in the installation (EBELING, 1997).

The maintenance delay time is the time spent waiting for maintenance facilities or resources. We consider a team that performs specialized in each data center that performs both preventive and corrective actions.

5.3.1 Input Parameters

The input parameters for the SPN model can be related to values or guard expressions. States receive the number of tokens that are triggered in the face of a timed or immediate transition. For our SPN model, this value can be zero or one. We consider the values in hours to represent MTTFs, MTTRs, administrative time, and the time interval between preventive maintenance for timed transitions. The immediate transitions received different guard expressions, depending on what it wants to verify. Table 32 presents a description of the states, timed transitions, and the SPN model's immediate transitions, composed of two subnets. The first represents the data center's maintenance actions, and the second checks the status of the system, checking whether the data center is active or inactive. Additionally, in the last column, the possible values or guard expressions for each state or transition are presented.

5.3.2 Proposed SPN Model

From the description presented in Table 32, it is possible to understand how we represent our maintenance policy for each state and transition. Figure 32 presents the SPN model proposed, which consists of two sub-nets representing and verifying the data center states, left network and right network.

The left network has the states DC_Up and DC_Down and the maintenance actions. The network on the right has the states $SystemUP$ and $SystemDown$ and checks the data center's general condition. The sub-net suggests the need for preventive repair from the guard expression $(\#DC\_up = 0)$ in the $FailureDC$ immediate transition, and a token is placed in the $SystemDown$ place. Suppose the team performed the repair, the sub-net, from the guard

Table 32 – Input Parameters for SPN Model

#	Name	Description	Value or Guard Expression
Places	DC_up	The mark assigned to the DC_up place indicates the data center is in an operational state.	0 or 1
	P3	Represent an Erlang distribution to indicate critical moments, which need repairs but did not cause the DC to stop.	
	P4		
	P5		
	DC_down	If the repair is not carried out, a mark will be assigned to the location DC_down, indicating the unavailability of the data center as well as the need for corrective repair.	
	CM_Team	Activates the team to perform corrective maintenance.	
	RepairCM	Corrective repair performe.	
	Team	The team that carries out preventive and corrective maintenance. The token assigned to the $Team$ place represents the availability of its.	
	RepairPM	Preventive repair.	
Timed Transitions	mttf1	Receive the tier MTTF.	8,137 or 5,337 or 3,334 or 1,640
	mttf2		
	AdmTime	Administrative time to carry out the procedures for activating maintenance, calling up the team, and separating replacement components.	2
	tb_PM	Recieves interval between preventive maintenance achievements.	720 or 360 or 168
	mttr_PM	Receive MTTR for preventive maintenance.	12 or 8 or 4
	mttr_CM	Receive MTTR for corrective maintenance.	
Immediate Transitions	TI2	Enables next critical state.	Default
	TI3	Indicates need for preventive repair	(#RepairPM>0)
	TI4		(#RepairPM>0)
	TI5		(#RepairPM>0)
	TI6	Enables next fall state.	Default
	CM	Enables corrective repair.	
System Subnet			
Places	SystemUP	Indicates that the data is active.	0 or 1
	SystemDown	Indicates that the data is inactive.	
Immediate Transitions	FailureDC	Indicates data center failure.	(#DC_up=0)
	RepairDC	Indicates that the data center has been repaired.	(#DC_up=1)

Source – Elaborated by the author.

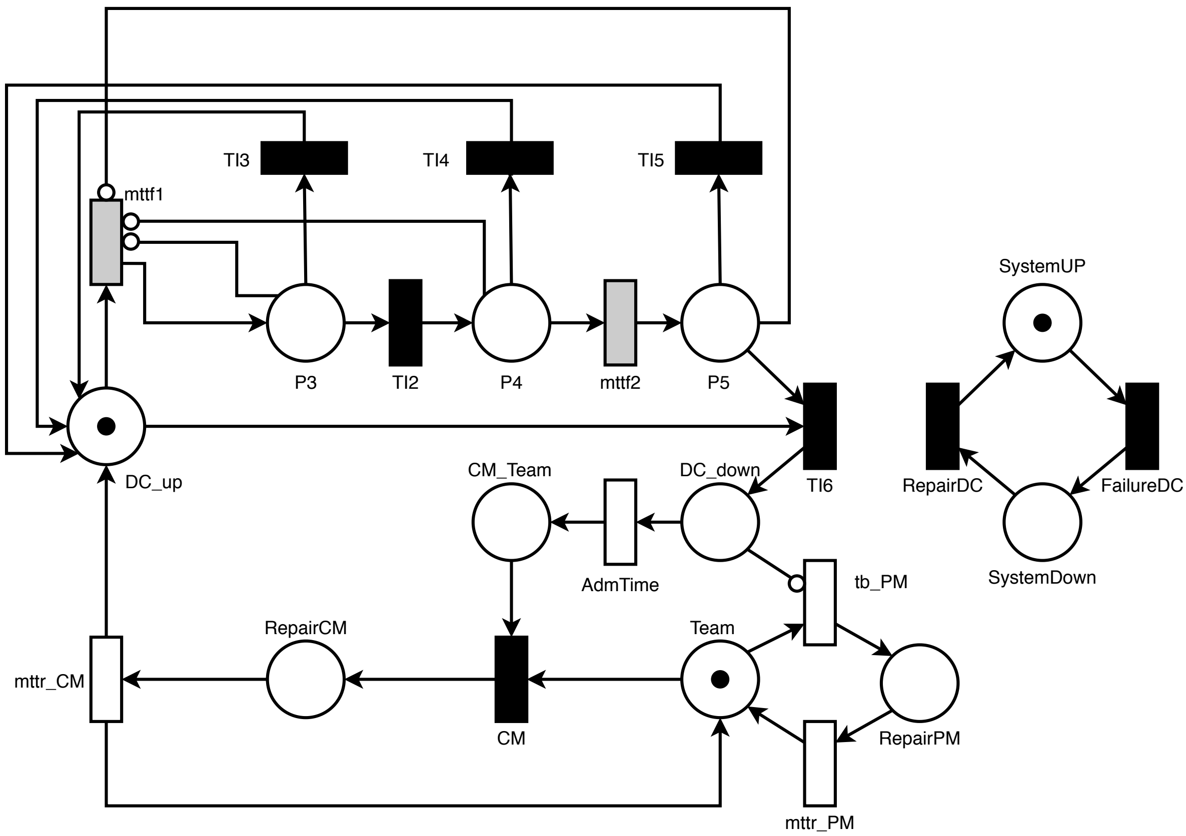
expression $(\#DC\_up = 1)$ in the $RepairDC$ immediate transition, checks if the system has been restored. If so, a mark is assigned to the $SystemUP$ place again, restoring the system to regular operation. Thus, in the data center network (left network), a token returns to the DC_Up state, indicating that everything is occurring under normal conditions.

If there is a token deposited in one of the states P3, P4 and P5, there is a need to perform a preventive repair on some component of the network, but that situation has not yet led to the data center failure. However, if the preventive repair is not successful, the network will go to an inactive state (DC_Down), where there will be an administrative time of two hours to activate the corrective maintenance team ($Team$), and after the corrections ($Repair\_CM$) the data center returns to operational status again, and the team is released.

Note that in the event of a data center failure, only corrective maintenance can be performed. Moreover, if preventive maintenance is being carried out, activated by transitions TI1, TI4, TI5, the token is removed from the $Team$ place, and corrective maintenance is not possible.

The inhibitory arc, which links a given state to a transition, serves to inhibit the transition's triggering if a mark is deposited in the place where the arc started. For example, the inhibiting arc that leaves the DC_down place prevents preventive maintenance from being performed after the system failure. That is, in this case, only corrective maintenance can be performed.

Figure 32 – SPN Model - Maintenance policy.



Source – Elaborated by the author.

5.4 FINAL CONSIDERATIONS

This chapter presented an efficient calculation for the demand for data center loads, the modeling for electrical and network architectures, and maintenance policy, including the EFM, RBD, and SPN models' heterogeneity. Definitions, input parameters, and proposed models were presented for each model category.

6 EXPERIMENTATION

This chapter presents model evaluations and experiments created in different scenarios for the electrical and IT architectures. Therefore, we divided the chapter into two sections, electrical architecture and IT architecture.

For electrical architecture, the evaluations were carried out to:

- verify the energy flow, quantify exergy and costs, sustainability metrics, and dependability metrics;
- calculate the environmental impact caused by different types of raw materials used for energy production, from different scenarios; and
- identify the main bottlenecks in architectures from the application of the parametric sensitivity analysis technique.

For IT architecture, the evaluations were performed to:

- verify dependability metrics such as reliability, availability, and mean time to failure;
- identify the main bottlenecks in architectures from the application of the parametric sensitivity analysis technique; and
- identify the impact of a maintenance policy on the availability of data center networks.

Thus, we were able to identify several important points that deserve more attention during the planning phase of the data center designs and suggest improvements to the designed architectures.

The analyzes consider variations in the period between evaluations, MTTFs, and MTTRs.

Scenarios for monitoring parameters are presented as proposed in SLA contracts to match actual estimates.

The evaluations' results allow the management of service levels and provide evidence about the models' usefulness and effectiveness and the approach proposed in this thesis.

6.1 ELECTRICAL ARCHITECTURE

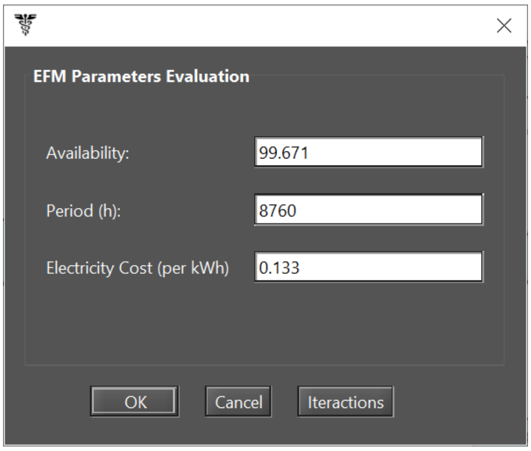
The assessment of an electrical system's reliability should include a review of the system at several levels, such as utility supply, configuration, control and protection, physical installation, operations, and maintenance.

Based on the evaluations, it is possible to present better configurations for the electrical system's availability to allow effective planning and physical arrangements that allow the data center's growth without changing the existing distribution paths. It is possible to identify whether the redundant paths are provided with physical segregation so that a severe failure of a piece of equipment cannot readily propagate to the redundant circuits or equipment. Identify appropriate periods for carrying out maintenance and the parts that deserve to have spare components to reduce repair times in the event of failure (HEISING et al., 2007). Among many other issues that can be better analyzed in the design phase, results are comparable to real situations.

In this thesis, we executed several evaluations based on the models represented to identify whether, in the way the architectures were designed, it is possible to achieve specified service levels according to the tier classification, being the most significant point of comparison the availability and downtime. Also, we present other metrics of dependability, sustainability, and costs. We performed several experiments to calculate how much CO_2 a data center can emit when choosing the primary energy production source. We identified the possible points of failure of the represented architectures using the parametric sensitivity analysis technique. Throughout the chapter, the evaluations, and experiments for each type of architecture (electrical and IT) are detailed.

As mentioned, we used the Mercury Tool (MACIEL et al., 2017) to generate models and evaluations in the three modeling techniques (EFM, RBD, and SPN). Figure 33 presents the Mercury interface corresponding to the input parameters for each EFM model's evaluation. For each tier, we inform the minimum availability recommended (see Table 33), performed evaluations for 720 hours (one month) and 8760 hours (one year), especially to highlight the difference between the values of total exergy and the total cost over the evaluated periods, and we consider the energy cost for all tiers of $ 0.133 (current cost of energy in the United States).

Figure 33 – Mercury Tool - EFM Parameters Evaluation.



Source – Mercury Tool (MACIEL et al., 2017)

Table 33 presents the suggested availability as well as the downtime, for each type of Tier (from 1 to 4), according to the classifications given in IV et al. (2006). We take the values in this table as a basis for making the comparisons after evaluating the availability achieved with each architecture represented for the tiers, be it electrical or IT architecture.

Table 33 – Availability and Downtime by Data Center Tier.

Tier	Avaiability (%)	Downtime (hr)
1	99.671	28.8
2	99.749	22.0
3	99.982	1.6
4	99.995	0.4

Source – Elaborated by the author.

6.1.1 Results of Evaluations

The EFM models representing the electrical architectures are shown in Figures 17, 19, 21 and 23. After evaluating each model, we obtained different metrics as presented in Table 34, which shows the embedded exergy (Em. Exe.) results, consumption of operational exergy (Op. Exergy), which can be understood as the fraction of the heat dissipated by each item of equipment that cannot be theoretically converted into useful work, and total exergy. Acquisition costs (Ac. Cost), operating costs (Op. Cost), and Total Costs (in American dollar) are also shown. The operating cost was calculated according to Equation 2.13. The penultimate column (Input Power) shows the minimum recommended demand after models' evaluation. According to their efficiency specification, this value corresponds to the minimum load provided by ACSource, including the energy losses suffered by the components. The last column (Sys. Eff.) shows the system efficiency. Note that the greater the redundancy (Tier 3 and 4), the less efficient the system is.

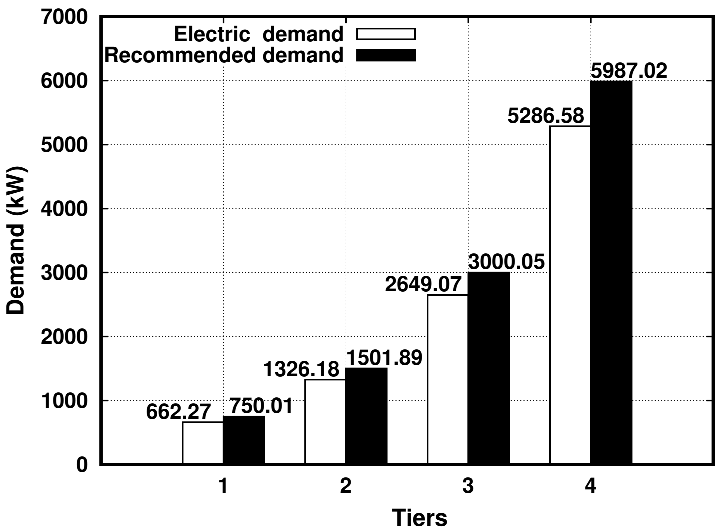
Table 34 – Results of Evaluations of Electrical Architectures.

Tier (Model)	Eval. Period (h)	Em. Exe. (kw)	Op. Exergy (kw)	Total Exergy (kw)	Ac. Cost (US$)	Op. Cost (US$)	Total Cost (US$)	Input Power (kw)	Sys. Eff. (%)
1 (Fig. 17)	720	260	1,214,690	1,214,951	45,842.5	17,223,315.37	17,269,157.87	1,804	36.70
	8760		14,778,731	14,778,992		209,550,337.03	209,596,179.53		
2 (Fig. 19)	720	265	2,434,310	2,434,575	81,345.0	34,516,532.30	34,597,877.30	3,613	36.70
	8760		29,617,442	29,617,708		419,951,143.03	420,032,488.03		
3 (Fig. 21)	720	280	6,468,242	6,468,523	87,727.5	82,569,873.59	82,657,601.09	8,624	30.72
	8760		78,696,954	78,697,235		1,004,600,128.64	1,004,687,856.14		
4 (Fig. 23)	720	445	12,909,962	12,910,407	129,117.5	164,801,165.57	164,930,283.07	17,210	30.72
	8760		157,071,208	157,071,654		2,005,080,847.73	2,005,209,965.23		

Source – Elaborated by the author.

Figure 34 shows the minimum recommended demand after models' evaluation. This value corresponds to the minimum load that goes out to supply the critical IT load. The calculated value for IT electrical demand was calculated as presented in Section 5.1.1. However, the recommended value considers a gap in the electrical charge to ensure no critical charge stress.

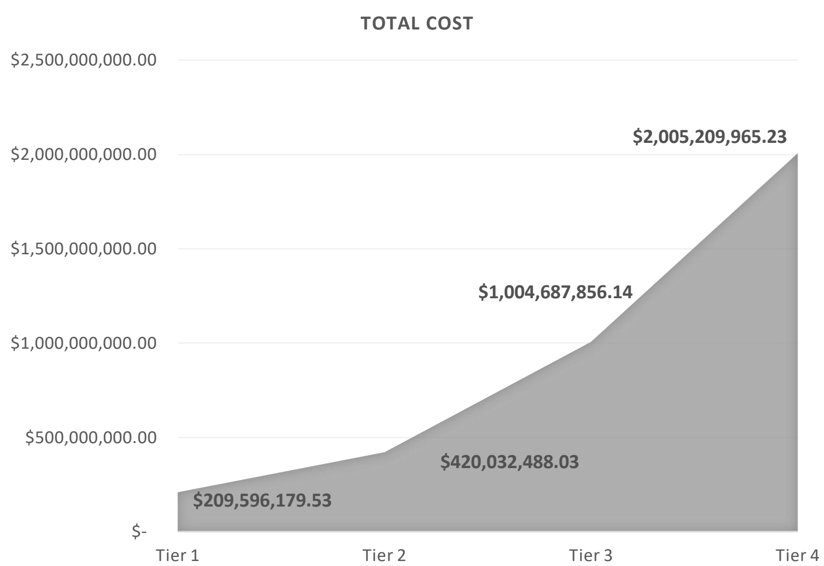
Figure 34 – DC Load in kW.



Source – Elaborated by the author.

Figure 35 shows the total cost of electrical architectures for each tier.

Figure 35 – Total Cost of Electrical Architectures.

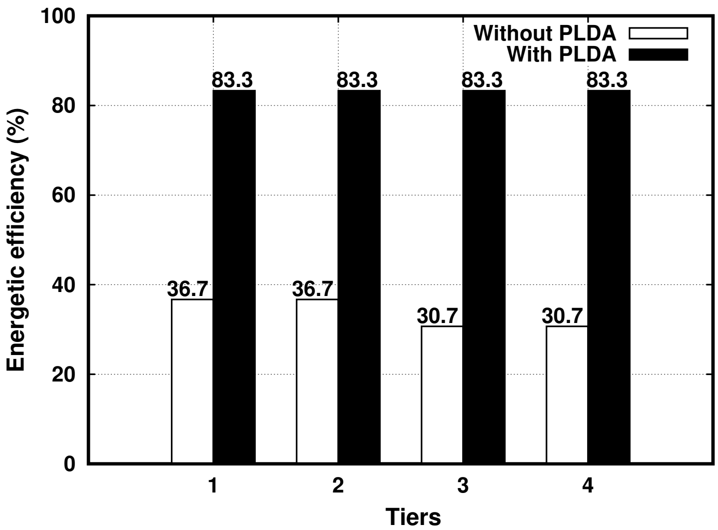


Source – Elaborated by the author.

After evaluating the models, a Power Load Distribution Algorithm (PLDA), based on the Ford-Fulkerson algorithm, was used to minimize the electric energy consumption of the models (MACIEL et al., 2017). The PLDA calculates the maximum possible flow, where the devices' carrying capacity is defined at the edges. The algorithm starts by going through the graph, looking for the best flows between two specific points on the graph. If a given path cannot support all the required flow, the residual flow will be redirected to other paths.

Figure 36 shows the gain in efficiency after executing the PLDA. This approach was able to achieve less consumption. With the PLDA, the efficiency of the four architectures rose to 88.3%. This result is achieved after the algorithm eliminates paths/components with low efficiency from the main energy flow. A practical example for the use of PLDA can be considered in an active-standby state, which, despite keeping the data center in operational mode, uses resources according to demand, which is especially important for lower-tier data centers (Tier 1 and 2) and energy-saving. Efficiency without PLDA can be seen in Table 34 (last column - Sys. Eff.).

Figure 36 – Efficiency Gain with PLDA.



Source – Elaborated by the author.

The dependability results achieved with the RBD models' evaluation scripts, performed to represent the electrical architecture, are presented in Table 35, whose metrics presented are availability, reliability, and MTTF. Observing the availability achieved and comparing the minimum suggested availability for each tier (see Table 33), we can conclude that all electrical architectures have a satisfactory level of redundancy to achieve the suggested availability.

Regarding reliability and MTTF, we can see that as redundancy increases, these metrics also increase; this is due to the addition of redundant components with high MTTF. Thus, for tier 3 and 4, the capacity to support failures is more pronounced, considering different paths.

Table 35 – Results of Evaluations – RBDs of Electrical Architectures.

Tier (Model)	**Evaluation Period (8760 h)**		
	Availability (%)	**Reliability (%)**	**MTTF System (h)**
1 (Fig. 17)	99.93398	7.88	3,891.37
2 (Fig. 19)	99.96088	9.87	4,275.98
3 (Fig. 21)	99.99996	15.13	5,452.59
4 (Fig. 23)	99.99998	24.90	6,806.54

Source – Elaborated by the author.

6.1.2 Experimental Studies

The experiments represent a situation provoked to observe, under control, the relationship between certain phenomena (GERHARDT, 2009). For this thesis, we carried out some experiments, which are represented in different scenarios, based on real information. Two electrical architecture experiments were conducted to show the use of different raw materials to calculate the environmental impact caused by CO_2 emissions into the atmosphere. To calculate the CO_2 emission, we use the Equation 2.14.

6.1.2.1 First Experiment

We chose three primary sources based on those widely used worldwide: coal, natural gas, and nuclear.

We also consider two sources of cleaner energy, with more excellent representation on the planet, such as wind and hydroelectric, whose aggression factors were obtained from Callou et al. (2013a), Ferreira et al. (2018). We consider diesel as the material used in generators, whose aggression factor was obtained from Quaschning (2020b).

We defined the input power used for the calculation to be the result of the model's evaluation, which are different for each tier (second line in the Table 36). Within the (minimum) availability that must be achieved by the tier, a part is represented using the primary source and another using generator. So, we have the following scenarios (CAMBOIM et al., 2020a):

- Tier 1 - 79.671% of coal use and 20% of the generator;
- Tier 2 - 84.749% of natural gas use and 15% of the generator;
- Tier 3 - 89.982% of nuclear energy use and 10% of the generator;
- Tier 4 - 50.0% of hydroelectric energy use, 44.995% of wind energy, and 5% generator. Being this the only tier that has two primary sources of supply.

Table 36, specifically in the second column, shows how many grams of CO_2 (or aggression factor) are emitted per kilowatt-hour for a given type of source (given in the first column). The number of kilograms of CO_2 issued per kWh is shown in the bottom row.

Table 36 – CO_2 Emission - Experiment 1.

Parameters	**CO_2 Emission (g/kWh)**	**Tier 1**	**Tier 2**	**Tier 3**	**Tier 4**
Input Power (kW)	–	1,804.53	3,613.55	8,624.14	17,210.67
Availability (%)	–	99.671	99.749	99.982	99.995
Wind (g/kW)	10				0.44995
Coal (g/kW)	950	0.79671			
Natural gas (g/kW)	515		0.84749		
Nuclear (g/kW)	150			0.89982	
Hydroelectric (g/kW)	20				0.5
Diesel (g/kW)	270	0.2	0.15	0.1	0.05
CO_2 Emission (kg/kWh)	–	1,463.25	1,723.51	1,396.88	481.89

Source – Elaborated by the author.

Analyzing the results, we were able to identify the environmental impact, given the emission of the main greenhouse gas (CO_2).

From the percentage of use of a certain type of material, we have the following results:

- Tier 1 – Produces 1,463.25 kg/kWh. This data center will have emitted **0.012** tons of CO_2 into the atmosphere over a year.
- Tier 2 – Produces 1,723.51 kg/kWh. This data center will have emitted **0.015** tons of CO_2 into the atmosphere over a year.
- Tier 3 – Produces 1,369.88 kg/kWh. This data center will have emitted **0.012** tons of CO_2 into the atmosphere over a year.
- Tier 4 – Produces 481.89 kg/kWh. Over a year, this data center will have emitted **0.004** tons of CO_2 into the atmosphere.

Take the tier 4 data center, for example, which has two primary power sources and eight times more IT equipment than DC tier 1; both sources used are cleaner than all the other sources represented; the CO_2 emission of this infrastructure is almost four times lower than that of the tier 2 data center, which had a more significant environmental impact.

6.1.2.2 Second Experiment

To perform this second experiment, we chose another's primary sources used in energy production, considering that they are used in various parts of the world. For example, refinery gas, gasoline, and lignite (a type of brown coal, quite common in Australia and Texas, can be used in energy production). We use the same clean energy sources as in experiment 1, the same suggested availability and input power. For other sources, the aggression factor was obtained from Quaschning (2020b).

The other definitions of experiment 1 are valid for this one; however, we have the following scenarios:

- Tier 1 - 79.671% of lignite use and 20% of the generator, which fueled by gasoline;
- Tier 2 - 79.749% of refinery gas use and 20% of the generator, which fueled by gasoline;
- Tier 3 - 79.982% of hydroelectric use, 10% of the generator fueled by gasoline, and 10% of the generator fueled by diesel;
- Tier 4 - 49.995% of wind energy use, 30.0% of hydroelectric energy use, 10% of the generator fueled by gasoline, and 10% of the generator fueled by diesel. Being this the only tier that has two primary sources of supply.

Table 37 indicates the modeled scenarios.

Table 37 – CO_2 Emission - Experiment 2.

Parameters	CO_2 **Emission (g/kWh)**	**Tier 1**	**Tier 2**	**Tier 3**	**Tier 4**
Input Power (kW)	–	1,804.53	3,613.55	8,624.14	17,210.67
Availability (%)	–	99.671	99.749	99.982	99.995
Wind (g/kW)	10				0.49995
Hydroelectric (g/kW)	20			0.79982	0.3
Refinery gas (g/kW)	240		0.79749		
Gasoline (g/kW)	250	0.2	0.2	0.1	0.1
Diesel (g/kW)	270			0.1	0.1
Lignite (g/kW)	360	0.79671			
CO_2 **Emission (kg/kWh)**	–	607.79	872.30	586.41	1,075.66

Source – Elaborated by the author.

Note that even for this scenario's unclean sources, the aggression factor is less than for experiment 1. Thus, we have the following results:

Tier 1 - Produces 607.79 kg/kWh. This data center will have emitted **0.00532** tons of CO_2 into the atmosphere over a year.

Tier 2 - Produces 872.30 kg/kWh. This data center will have emitted **0.00764** tons of CO_2 into the atmosphere over a year.

Tier 3 - Produces 586.41 kg/kWh. This data center will have emitted **0.00513** tons of CO_2 into the atmosphere over a year.

Tier 4 - Produces 1075.66 kg/kWh. This data center will have emitted **0.00942** tons of CO_2 into the atmosphere over a year.

The data center tier 3 has an energy demand almost five times greater than the data center tier 1 and still emits less CO_2 than tier 1; this is the same as saying that a single data center, which makes a wrong choice for its power source, issues the amount of CO_2 that five data centers use cleaner energy source, in this case, the hydroelectric.

6.1.2.3 Third Experiment

To perform the third and last electrical architecture experiment, we applied the parametric sensitivity analysis technique. We varied the component MTTFs and MTTRs to +50% and -50% of the original value (as shown in the last two columns of Table 28). As for all electrical architecture components, the MTTR was 8 hours, so the variation was from 4 to 12 hours. With this margin, we were able to identify the component that has the most significant impact on the system's availability.

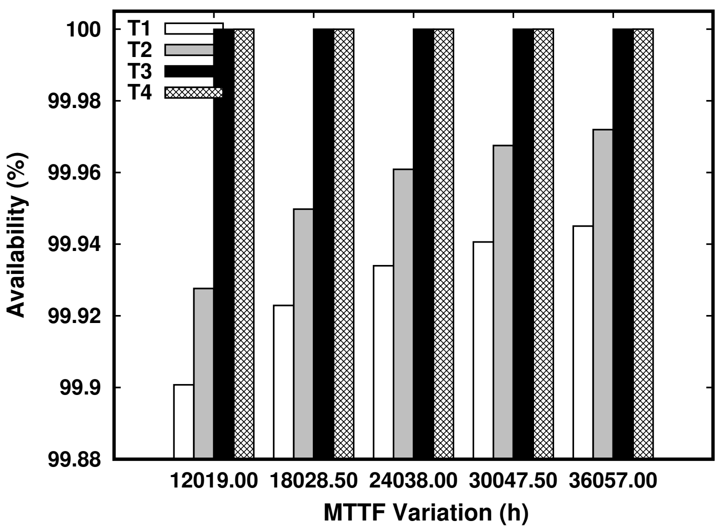
Table 38 shows the rank for the sensitivity analysis of the first five most important components for each tier. The sensitivity value and the reboot component are also displayed. The reboot component will have little or no influence on overall availability (its value is displayed in the same analyzes, being given as the last component of the ranking). Figure 37 shows the availability achieved according to the ATS components' MTTF variation. The highest sensitivity for tiers 1 and 2 is related to the MTTF variation of the ATS component, and for tiers 3 and 4, it is the MTTF of ATS1 (changing only the nomenclature due to the addition of redundancy).

Table 38 – Sensitivity Ranking of Electrical Architecture.

Model	Component	Rank	Value	Reboot
RBD-EFM T1DC (Fig. 18)	mttf_ATS	1st	4.4E-04	mttr_Jb$_i$
	mttr_ATS	2nd	3.3E-04	
	mttf_UPS1	3rd	2.1E-04	
	mttr_UPS1	4th	1.6E-04	
	mttf_SDT1	5th	7.5E-05	
RBD-EFM T2DC (Fig. 20)	mttf_ATS	1st	4.4E-04	mttr_PS$_i$
	mttr_ATS	2nd	3.3E-04	
	mttf_SubP1	3rd	7.0E-05	
	mttr_SubP1	4th	5.3E-05	
	mttf_ACS	5th	7.7E-06	
RBD-EFM T3DC (Fig. 22)	mttf_ATS1	1st	2.9E-07	mttr_Jb$_i$
	mttr_ATS1	2nd	2.2E-07	
	mttf_ATS2	3rd	1.7E-07	
	mttr_ATS2	4th	1.3E-07	
	mttf_UPS3	5th	8.2E-08	
RBD-EFM T4DC (Fig. 24)	mttf_ATS1	1st	2.2E-07	mttr_PS$_i$
	mttf_ATS2	2nd	1.7E-07	
	mttr_ATS1	3rd	1.6E-07	
	mttr_ATS2	4th	1.3E-07	
	mttf_SubP1	5th	3.5E-08	

Source – Elaborated by the author.

Figure 37 – Availability of ATS Components.



Source – Elaborated by the author

The lower the MTTF value of these components, the greater the system downtime. Thus, we can conclude that this component is the most important for the four represented architectures' availability. In this sense, we can request more attention regarding maintenance and repair and instruct the need for redundancy and spare pieces.

6.1.3 Discussion of Results

The results indicate that as higher the tier classification, the lower energy efficiency, and the higher the costs and energy consumption; this is due to the need to feed more alternative systems/paths (from tier 1 to 4) without necessarily having useful work on all paths, to standby for the case of failure of any component or system.

The represented electrical architectures all manage to reach the minimum level of availability suggested according to their classification. We present the results for exergy and costs of the architectures, and in the way presented, it is possible to observe how expensive it is to maintain the sites and their operations when observed for one month and one year. Using the flow optimization algorithm (PLDA), we could obtain better efficiency for the four represented architectures.

From the results obtained with the evaluations of the RBDs models, we were able to identify that the electrical architectures proposed for this work manage to reach the minimum level of availability suggested, given the classification by tier. However, the more sophisticated the data center, the higher its operating costs.

We created some scenarios to represent different raw material types to serve energy production sources from the input power resulting from the EFM modeling. In this way, we evaluate the environmental impact caused by emissions of CO_2 into the atmosphere.

We conclude that the worst impact on environmental sustainability does not come from the DC layer classification but from the type of source that feeds it. The choice of raw material is much more essential and can negatively impact the environment than the load required to power the data center. Our experimental studies exemplified that tier 3 e 4 DCs can produce less CO_2 than a DC with less IT load (least tier). It is worth highlighting the importance of using cleaner energy sources, those that produce less CO_2.

Research shows that the dealership supply is the largest individual component that affects an industrial plant (HEISING et al., 2007). We were able to identify which component has the most significant impact on the electrical system's availability of the modeled architectures from sensitivity analysis. Thus, we can suggest some proposals that can decrease downtime.

> Response to partial load conditions is important to understand when estimating overall energy consumption in a data center with varying IT loads. Also, while multiple concurrently energized power distribution paths can increase the availability (reliability) of the operations, this type of topology can decrease the overall system's efficiency, especially at partial IT loads. (GENG, 2014, p. 32).

6.2 IT ARCHITECTURE

To represent the four tiers' components and IT architectures, we created seven RBDs models (as presented in Section 5.2). The first three models represent the access layer components, and the subsequent four models represent the tier networks (Tier 1 to 4). All architectural sensitivity analysis were performed using the Mercury Tool scripting language (MACIEL et al., 2017). Listing 6.1 shows the 8760-hour evaluation script example of the first RBD network model, the blade server (RBD-Net - Server, Fig. 25). The other scripts for electrical and network architectures are presented in Appendix C.

Listing 6.1 – Mercury Script for Blade Server RBD Evaluation.

```
t = 8760;
RBD Model{
   block BaseBlade( MTTF = 220000.0, MTTR = 4.0);
   block CPU( MTTF = 50000.0, MTTR = 4.0);
   block DIMMs( MTTF = 480000.0, MTTR = 4.0);
   block HD( MTTF = 200000, MTTR = 4.0);
   block FiberChanelCard( MTTF = 260000.0, MTTR = 4.0);
   block EthernetCard( MTTF = 1240000.0, MTTR = 4.0);
   series s0(BaseBlade, CPU, DIMMs, HD, FiberChanelCard, EthernetCard );
   top s0;
   metric av = availability;
   metric rel = reliability( time = t );
   metric mttf = mttf;
   metric mttr = mttr;
} main{
  av = solve(Model, av);
  rel = solve(Model, rel);
  mttf = solve(Model, mttf);
  mttr = solve(Model, mttr);
  println("Availability: " .. av );
  println("Reliability: " .. rel );
  println("Mean time to failure: " .. mttf );
  println("Mean time to repair: " .. mttr );}
```

From the results obtained with the networks' RBDs models' evaluations, we created different experimentation scenarios to analyze the architecture's availability according to the proposed maintenance policies (SPN model), which corresponds to nine SLA classes.

6.2.1 Results of Evaluations

Table 39 shows the evaluation results of the seven RBD models (Chapter 5). The evaluation periods correspond to 720 and 8760 hours. The focus for choosing these times is to show how reliability behaves. System reliability is directly related to MTTF parameters, and the system availability is directly related to MTTF and MTTR parameters (KUO; ZUO, 2003). Note that availability and MTTF are the same for both periods. However, for reliability, specifically for the network models (last four - Tier 1 to 4), the measure that we increased the components redundancy and achieved greater availability, the system's reliability decreases considerably. Likewise, the system's mean time to failure is shorter. In this sense, if we consider a tier 4 DC, which should provide a minimum availability of 99.995%, compared to a tier 1 DC, preventive maintenance must be carried out more frequently to ensure the quality of service at a high confidence level; this happens due to the need to maintain more components in simultaneous operation.

Table 39 – Results of Evaluations – RBDs of IT Architectures.

Model	**Evaluation Period (h)**	**Reliability (%)**	**Availability (%)**	**MTTF (h)**
RBD-Net - Server (Fig. 25)	720	97.42	99.98548	27,562
	8760	72.77		
RBD-Net - Chassis (Fig. 26)	720	95.74	99.97588	12,634
	8760	58.37		
RBD-Net - Rack (Fig. 27)	720	99.99	1	28,849
	8760	96.87		
RBD-Net - T1DC (Fig. 28)	720	98.38	99.98226	16,274
	8760	80.11		
RBD-Net - T2DC (Fig. 29)	720	99.24	99.99168	10,675
	8760	68.57		
RBD-Net - T3DC (Fig. 30)	720	99.23	99.99167	6,668
	8760	15.42		
RBD-Net - T4DC (Fig. 31)	720	99.95	99.99999	3,281
	8760	8.47E-06		

Source – Elaborated by the author.

6.2.2 Experimental Studies

For the IT architecture, we performed two experiments. The first corresponds to the sensitivity analysis of the network components, and the second corresponds to the evaluation of the maintenance policy.

6.2.2.1 First Experiment

We applied the differential parametric sensitivity analysis for all RBD models, except for the third, given that it presents only components of equal importance, to identify the essential components, and so, establish prioritization actions concerning their maintenance and replacement.

As with electrical architectures, we also vary the MTTF and MTTR of each component between -50% to +50% of the original value to perform the sensitivity analysis. The MTTF variation is shown in the last two columns of Tables 29, 30, 31. For components with a 4-hour MTTR, we range from 2 to 6, and with an MTTR of 8, from 4 to 12. All scripts for the parametric sensitivity analysis are presented in Appendix C.

Table 40 shows the sensitivity index's rank, with the five most important components for availability, the sensitivity value, and the component that will have the least significant impact on the system's availability (reboot component). In most cases, the most critical elements for availability are also more essential for reliability.

Figure 38 shows the impact of the essential components on system availability. For all models, MTTF variation had a more significant impact than the MTTR. All graphs were plotted using the same scale of availability (y-axis), ranging from 99.95% to 100%. Note that the way the networks were represented, they all reach the availability that is suggested for each tier (see Table 33).

For the first and second models (RBD-Net - Server and RBD-Net - Chassis), the MTTF of the CPU and software have the most significant impact on the blade center system (Figure 38a and 38b). Considering that the software is the component that has the lowest original MTTF presented (two years - see Table 30), it is precisely in the second model that we find the least availability.

For the network graphs (Figure 38c to 38f), we can see the difference between the variation in availability, especially between tiers 1 and 4 (Figure 38c and 38f). For tier 1, even raising the MTTF of the switch to +50% of the original value, the availability does not exceed 99.985%; this is due to the absence of component redundancy. For tier 4, which has the highest level of redundancy between components, availability is the highest.

6.2.2.2 Second Experiment

For the evaluation of maintenance policies, we will only use tier data center networks. In this way, we will be able to analyze the variation in availability under the proposed policy. The MTTFs achieved by analyzing these models are used as input metrics in the SPN model. In this sense, we evaluate the SPN model often to find the availability achieved in each network and observe the maintenance effects in the network's availability.

The SPN model was designed to represent the proposed maintenance policy. We consider the intervals between these maintenances, which we classify as "poorly managed" - when a maintenance occurs every 30 days (720 hours), "managed" - every 15 days (360 hours), and the "well managed" - every seven days (168 hours). Also, we adopt different repair times for each category, varying between -50% and + 50% of the original MTTR (8 hours – from 4 to 12).

Table 40 – Sensitivity Ranking of IT Architecture.

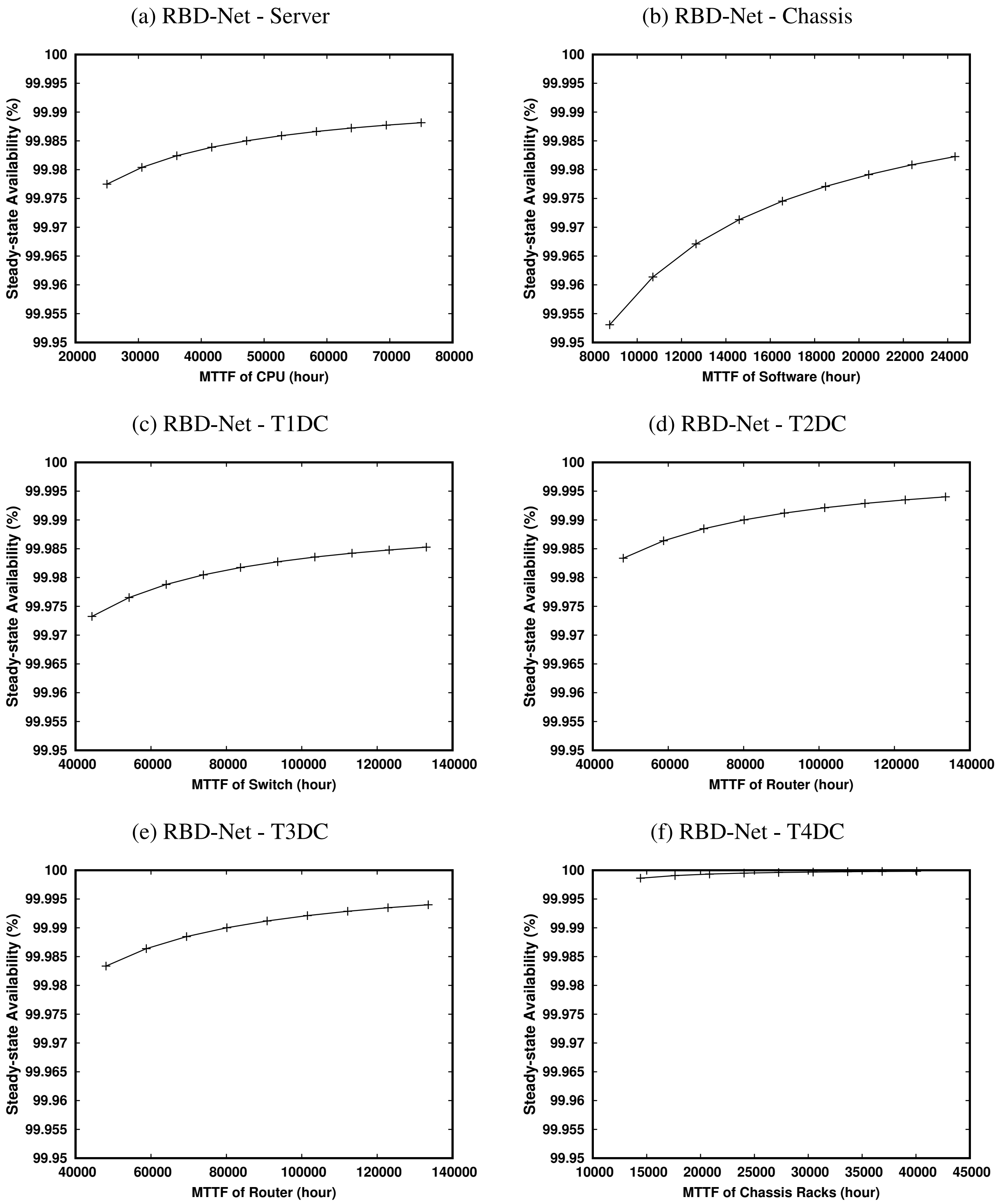
Model	Component	Rank	Value	Reboot
RBD-Net - Server (Fig. 25)	mttf_CPU	1st	1.07E-04	mttr_EthernetCard
	mttr_CPU	2nd	8.00E-05	
	mttf_HD	3rd	2.67E-05	
	mttf_BaseBlade	4th	2.42E-05	
	mttf_FiberCard	5th	2.05E-05	
RBD-Net - Chassis (Fig. 26)	mttf_Software	1st	3.04E-04	mttr_B_Server
	mttr_Software	2nd	2.28E-04	
	mttf_Midplane	3rd	1.72E-05	
	mttr_Midplane	4th	1.29E-05	
	mttf_EtherSwitch	5th	3.95E-09	
RBD-Net - Rack (Fig. 27)	–	–	–	–
RBD-Net - T1DC (Fig. 28)	mttf_SW	1st	1.20E-04	mttr_CR
	mttf_RT	2nd	1.11E-04	
	mttr_SW	3rd	9.02E-05	
	mttr_RT	4th	8.32E-05	
	mttf_QSFP	5th	5.26E-06	
RBD-Net - T2DC (Fig. 29)	mttf_RT	1st	1.06E-04	mttr_SFP
	mttr_RT	2nd	7.39E-05	
	mttf_SW2	3rd	1.08E-08	
	mttf_SW	4th	7.23E-09	
	mttr_SW	5th	7.23E-09	
RBD-Net - T3DC (Fig. 30)	mttf_RT	1st	1.11E-04	mttr_UTP
	mttr_RT	2nd	8.32E-05	
	mttf_SW	3rd	8.68E-08	
	mttr_SW	4th	4.88E-08	
	mttf_CR	5th	1.96E-08	
RBD-Net - T4DC (Fig. 31)	mttf_CR	1st	1.23E-05	mttr_SFP
	mttr_CR	2nd	6.90E-06	
	mttf_SW	3rd	1.73E-07	
	mttr_SW	4th	9.76E-08	
	mttf_RT2	5th	9.23E-09	

Source – Elaborated by the author.

We consider nine scenarios with different service levels, which can be found in SLA contracts, to check if the intervals and repair times determined for the maintenance are sufficient for these networks to achieve the minimum availability suggested for the tier. Remembering that if a failure could cause the system to down, only corrective maintenance can be performed to correct it and make the system work again.

Figure 39 presents the Mercury interface to perform the SPN model's stationary simulation. We chose the 99% confidence level and the maximum relative error of 5%. For the other parameters, we leave the default value.

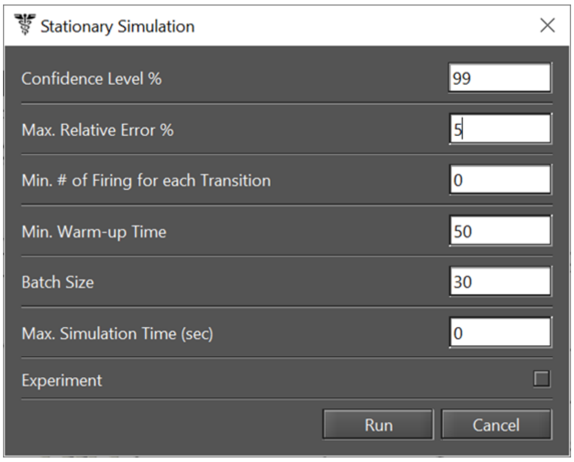
Figure 38 – Effect of Each 1st Rank Component on Systems' Availability



Source – Elaborated by the author.

Table 41 presents nine SLA classes that aim to represent the different management categories. The results presented for the nine SLA classes show a variation in the availability value achieved for each scenario, which we decided to compare with the availability suggested for each tier.

Figure 39 – Mercury Tool - Stationary Simulation.



Source – Mercury Tool (MACIEL et al., 2017)

Table 41 – Availability Achieved with the SPN Model.

SLAs	**Interval (h)**	**Repair (h)**	**Availability (%)**			
			Tier 1	**Tier 2**	**Tier 3**	**Tier 4**
1		12	91.60	88.16	83.09	71.24
2	720	8	91.88	89.52	82.35	70.73
3		4	92.50	87.09	81.49	69.98
4		12	96.45	93.78	90.62	82.28
5	360	8	96.10	93.45	90.22	82.67
6		4	96.95	92.42	90.29	82.15
7		12	98.10	97.57	94.48	91.82
8	168	8	97.86	96.83	95.70	91.74
9		4	98.22	96.93	95.35	91.08

Source – Elaborated by the author.

Some statements can be stated. The first is that we can infer that with the MTTFs of the components used, considering the proposed maintenance policy, does not make it possible to achieve high availability, as suggested for high-performance data centers. The highest achieved availability was that of SLA 9, tier 1, with 98.22%. If we compare with the availability indicated for this tier, that is 99.67%; no SLA can achieve it.

For tiers 1, 2, and 3, that despite being complete networks, still presenting a relatively low number of components; the levels of availability are merely unacceptable as a result of the lack of redundancy of the main components, such as the edge router, which was not represented due to the lack of redundancy of Internet providers. When the intervals between preventive maintenance are longer (e.g., 720 hours), availability has the worst result in all networks (1, 2, and 3 SLAs).

Note especially tier 4, which should provide availability with at least four nines, despite being a network that has all redundant components, and it has a very low mean time to failure, which greatly influenced the result. It is also notable that due to the number of components, it is unacceptable that preventive maintenance occurrences and repair times are as shown.

6.2.3 Discussion of Results

For the IT architecture of the networks and the dependability assessments, we carried out two experiments to present the essential components and the influence of specific metrics on the system's availability under a maintenance policy.

With the sensitivity analysis, we were able to identify which components are most significant for each architecture. With the evaluation of the maintenance policy, our goal was to meet nine different SLAs that were also proposed to meet the needs of tier 1 through 4 data centers. The results showed the difficulty of achieving high availability when the system's MTTFs are relatively low and when preventive maintenance is dependent on long intervals.

For operational guarantees, the data center design must provide maintenance policies whose intervals are sufficient to avoid service disruption. In situations where the downtime is more significant than specified in the SLAs contracts, the company offering the service would have to pay fines for non-compliance.

We could have represented other alternatives to provide higher availability results. The first would be to have shorter intervals between maintenance. The second would be to have lower MTTRs. Furthermore, the third would be to use higher MTTFs. However, we decided to leave the values as planned to show how availability is influenced by several parameters and how maintenance, especially preventive ones, can help maintain the system's operational status, avoiding the shortage of services.

Thus, we highlight the importance of short intervals between preventive maintenance and how much time spent on repairs plays a crucial role in ensuring availability. In general, switches and routers were presented as the essential components for networks; therefore, some practices aimed at more excellent reliability of these components would mitigate unplanned downtime risks.

Preventive maintenance usually receives little emphasis on the design and operation of distribution systems; however, it is critical in improving availability. Large expenditure on systems is made to provide the desired reliability; however, failure to provide timely and high-quality preventive maintenance measures leads to system or component malfunction or failure and prevents the intended design objective from being achieved. When an effective preventive maintenance program covers equipment, it performs better and lasts longer, so it is possible to reduce errors, accidents, and problems in advance, minimizing unplanned downtime, which is expensive (HEISING et al., 2007, p. 105).

The system design must be influenced, considering the initial installation cost, considering the maintenance actions. In most cases, the additional cost of performing maintenance, significantly no more extended schedule, lost production due to interruptions due to lack of maintenance, more than compensates for the initial cost savings (HEISING et al., 2007, p. 109).

6.3 FINAL CONSIDERATIONS

This chapter presented the experiments and the evaluations of the models of the electrical and network architectures. As a result, different metrics of dependability and sustainability were presented. Besides, for each type of architecture, the results were discussed. We vary the MTTFs and MTTRs of the architecture components to +50% and -50% for the sensitivity analysis. To calculate the environmental impact of using energy from large data center infrastructures, we consider the amount of CO_2 emissions according to the type of raw material used in energy production. We created different categories to analyze the SLA classes, varying the intervals between preventive maintenance and different repair times, following the proposed policy to investigate networks' availability.

7 FINAL REMARKS

Faced with the demand to maintain the high availability of data centers (DC), companies are being pressured to seek sustainable alternatives, given that these infrastructures consume a total of 1% of all electricity worldwide (SVERDLIK, 2020). In a time of pandemic (COVID-19), when the digital economy has assumed an even greater share of representativeness, DCs and telecommunications companies need to meet the requisitions of "everything-as-a-service". Linked to this are the large amounts of carbon dioxide (CO_2) emitted into the atmosphere due to the production and consumption of energy caused by these infrastructures. Besides, the design, implementation, and maintenance of data center networks must meet numerous dependability requirements to guarantee the quality of service at a high level of reliability and high availability.

Given the above, this work proposes models based on EFM, RBD, and SPN to represent both electrical and network architectures of data centers for tiers 1, 2, 3, and 4, to quantify and evaluate different dependability and sustainability metrics, and costs, especially since most related works deal with few characteristics presented in this thesis and consider only one type of architecture for data centers (or private cloud infrastructure).

From the input power resulting from the EFM modeling, we created some scenarios to represent different raw material types to serve as sources for energy production. In this way, we evaluate the environmental impact caused by emissions of CO_2 into the atmosphere.

The results achieved in this work indicate that for electrical architectures, in the way it was designed, it is possible to achieve the minimum availability suggested for each tier in the same way as network architectures. Besides, we can infer that the higher the tier classification, we had a lower result for energy efficiency, in addition to the fact that costs and energy consumption were higher than tiers 1 and 2; this is due to the redundancy of components for alternative paths for the highest-ranked tiers.

We conclude that the choice of raw material used in energy production that powers large data centers is much more essential and can negatively impact the environment than the load required to power the data center. Given the scenarios presented and the comparisons made, we prove the importance of making cleaner alternatives to promote environmental sustainability. It is worth mentioning that energy production and consumption will directly impact global warming, causing severe climate changes.

From the sensitivity analysis of the RBD models of both electrical and IT architectures, it was possible to identify the most critical components for the architectures' availability represented in this work. Thus, we suggest some proposals that can reduce the downtime.

To keep the data center networks in operational mode, we aim to comply with nine different SLAs proposed to meet the needs of tiers in the data center, where there is minimal redundancy of Internet providers and energy operators. However, according to the policy proposed in the SPN model, the approach proved to be insufficient to guarantee the high availability required for the tiers and, therefore, an approach that considers shorter preventive maintenance intervals may be considered to ensure the suggested minimum availability.

The results showed the difficulty of achieving high availability when the system's MTTFs are relatively low and when preventive maintenance is dependent on long intervals and demonstrate an increase in availability when preventive maintenance is performed more frequently. So, we emphasize that the simple fact that preventive maintenance reduces the risk of unscheduled downtime due to component failures already makes it a better option than corrective maintenance.

It is worth mentioning that the SLA management allows the administration of indicators and verifies problems, allowing proactivity that will bring benefits to the parties involved in the contract. Besides, a well-made SLA contract attests to the service provider's quality and can still work as a competitive advantage.

To promote the adoption of best practices regarding the quality, sustainability, and adherence of data center projects (and cloud infrastructures), a set of techniques and standards, as well as their definitions, were presented to guide companies and users regarding the implementation use of IT resources.

7.1 CONTRIBUTIONS

This thesis presented a set of models for representing and analyzing different data center architectures according to redundancy specifications by tier. Thus, we can consider issues of dependability, sustainability, costs, and sensitivity analysis. The adopted methodology proposes a hybrid and hierarchical approach so that the infrastructures are evaluated individually and allow the composition of simpler models. Therefore, we have as main contributions:

- The blade center system represents the choice of more efficient equipment for the data centers' access layer composition. In (DAMASCENO, 2015), (Smith et al., 2008), and (LEIGH; RANGANATHAN, 2007), the characteristics of efficiency and density make this system the best option available for current projects. Specifically, in (DAMASCENO, 2015), the author presents a set of comparisons between this and other systems that prove the best cost-benefit and return on the blade structures systems' investment. This way, we can promote energy and resource savings in data centers, consequently contributing to less CO_2 emissions.

- The methodology presented is based on an integrated proposal for a more efficient calculation of the load demanded for data center power architectures and for choosing more efficient components, both for electrical and network architecture. Besides, the methodology allows the generation of hybrid models and the use of metrics hierarchically so that it is possible to represent and evaluate different architectures, which were classified based on the component redundancy characteristics by tier. The experiments lead to suggestions for architectural improvements that can assist in efficient and sustainable infrastructure planning.

- The proposed modeling comprises three different types of techniques: energy flow models (EFM), reliability block diagrams (RBD), and stochastic Petri nets (SPN) for the representation of both electrical and IT architectures and for the proposition of a maintenance policy that includes preventive and corrective actions for data center networks. The models allow evaluating the sustainability, dependability, and maintainability requirements of the represented architectures and identify possible points of failure from the sensitivity analysis.

- Given the scenarios presented for electrical architectures, it is possible to quantify a specific type of raw material's environmental impact, generating energy. As each type of material has different aggression factors, we were able to identify different choices for less CO_2 emission. Therefore, the gap for promoting sustainability is reduced.

- All the architectures represented were focused on achieving energy efficiency (starting from the choice of components) and high availability (unanswerable requirements for data center). In this sense, there was a great effort to align the tiers' architectures according to the redundancy specifications. However, all the planning was thought to promote optimization of the use of resources and cost savings. In this way, we corroborate for planning of electrical and IT architectures for data centers through the physical redundancy of the components and still achieving dependability, sustainability, and cost requirements.

7.2 PUBLICATIONS

In addition to the contributions cited, the results of this work have generated some publications, which are highlighted next:

- IEEE SYSCON 2021
  Camboim, Kádna.; Ferreira, João.; Melo, Carlos; Araujo, Jean.; and Alencar, Fernanda. "Dependability and Sustainability Evaluation of Data Center Electrical Architectures". In: 15th Annual IEEE International Systems Conference (IEEE-SYSCON-2021).

- IEEE CloudNet 2020
  Camboim, Kádna.; Ferreira, João.; Araujo, Jean.; and Alencar, Fernanda. "Sustainability Analysis in Data Center Dense Architectures". In IEEE International Conference on Cloud Networking. 2020. (CAMBOIM et al., 2020a).

- SBESC 2020
  Camboim, Kádna.; Araujo, Jean.; Melo, Carlos; and Alencar, Fernanda. "Availability Evaluation and Maintenance Policy of Data Center Infrastructure". In Brazilian Symposium on Computing Systems Engineering (SBESC). 2020. (CAMBOIM et al., 2020b).

- Brazilian Journal of Development 2020
  Vale, Kádna Maria Alves Camboim, and Fernanda Maria Ribeiro de Alencar. "Challenges, Patterns and Sustainability Indicators for Cloud Computing." Brazilian Journal of Development 6, no. 8 (2020): 57031-57053. (VALE; ALENCAR, 2020)

- WER 2018
  Camboim, Kádna and Alencar, Fernanda. "*Requisitos não Funcionais e Sustentabilidade para Computação em Nuvem: uma Revisão Sistemática da Literatura.*" In 21st Workshop on Requirements Engineering. 2018. (CAMBOIM; ALENCAR, 2018)

- SBRC/WCGA 2018
  dos Santos, Danillo Moraes Lima, Kádna Maria Alves Camboim Vale, and Fernanda Maria Ribeiro de Alencar. "*Avaliação de desempenho de nuvens privadas: um comparativo entre Owncloud, Nextcloud e Pydio*. In: *Simpósio Brasileiro de Redes de Computadores. XVI Workshop em Clouds e Aplicações*, 2018, *Campos do Jordão - SP. Anais do XVI Workshop em Clouds e Aplicações*. 2018. v. 16. (SANTOS; CAMBOIM; ALENCAR, 2018)

7.3 LIMITATIONS

It is essential to highlight some limitations of this research:

- The first limitation is related to the researcher's possible biases, considering that a recommendation given by (KITCHENHAM; CHARTERS, 2007) is that at least two people carry out this process to minimize the researcher bias incorrect data extraction. A systematic research protocol was defined before its completion to mitigate this issue, and the process was followed from there.

- The second refers to the time and effort demanded regarding the realization of RSL given the excessive amount of papers returned from the search bases.

- The third limitation is that we try to be faithful to the MTTF values found for real components, but these values can differ significantly due to the variety of manufacturers' brands. Thus, our results do not represent this plurality.

- The fourth represents the minimum necessary of components to fit into the tier classifications. Still, in general, the actual data center infrastructure has many more features.

- The fifth is about the validation. Despite various experiments' performance, other techniques were not applied to validate the models and metrics presented.

- The last is the absence of the models representing the several redundancy mechanisms' logical connections.

7.4 FUTURE WORKS

Although this thesis presents a broad scope for an integrated solution for planning data centers, there are still margins for improvements and extensions, which are listed next as proposals for future work:

- We intend to extend the systematic literature review to extract detailed characteristics of the works related to this thesis and add works from snowballing to carry out a detailed analysis of the gaps that still exist in the context addressed.
- We aim to identify electrical costs based on the type of raw material chosen for energy production because, in this work, we use the same value for all EFM models, assuming the same value for all types of sources. In this way, we can evaluate the models several times and compare the costs. Also, we think to execute further experiments to identify CO_2 emissions using a single percentage for the use of primary sources by all tiers.
- We intend to carry out the cooling infrastructure modeling, which is the other primary subsystem of the data centers, to offer refrigeration efficiency mechanisms and contribute to sustainability.
- We expect to perform a sensitivity analysis with the "design of experiment" (DoE) for the critical components of the IT load so that it is possible to vary several metrics at a time and to compare several MTTF and MTTR values to expand the plurality concerning the supply, specifically MTTF, given by the manufacturers.
- We want to extend the SPN model to represent the inclusion of different maintenance teams and account for these teams' costs. Besides, we will consider real SLA contracts to compare the services contracted and achieved so that in the face of non-provision of the service as acquired, we must infer contractual fines. Another point that can be implemented is modeling based on existing real architecture, a functioning data center. Thus, to identify the most significant bottlenecks to suggest/make improvements, for example, we have the data center of this university.
- As we only represent physical architectures, we plan to create Markov chains and SPN models to simulate different active-standby and active-active logic redundancy protocols to the IT (network) architectures and compare their energy consumption. We also plan to compute the system's probability of entering a particular failure state to identify the architectures' most prominent points of failure.

7.5 SUMMARY

This thesis proposed a hybrid and hierarchical modeling approach for electrical and network architectures for data centers tier 1, 2, 3, and 4 to promote sustainability while achieving dependability requirements in planning these infrastructures. The contributions of this work revolve around the approach of methodology, modeling, evaluations, and experiments proposed. The thesis was structured in seven chapters, including the Introduction, Background, Methodology, Systematic Literature Review, Modeling, Experimentation, and Final Remarks. Appendices A, B, and C are presented after the Bibliography.

BIBLIOGRAPHY

ABDERRAHIM, W.; CHOUKAIR, Z. A framework architecture based model for cloud computing adaptive migration. In: **2014 Global Information Infrastructure and Networking Symposium (GIIS)**. [S.l.: s.n.], 2014. p. 1–6. ISSN 2150-3281. Cited on page 75.

ABNT. Nbr iso 14000 sistemas da gestão ambiental. 2005. Cited 2 times on pages 56 and 57.

ABNT. Sistemas da gestão ambiental, requisitos com orientações para uso. 2005. Cited on page 57.

ABNT. Nbr iso-iec 17788 tecnologia da informação computação em nuvem visão geral e vocabulário. Rio de Janeiro., 2015. Cited 2 times on pages 53 and 55.

AHSON, S. A.; ILYAS, M. **Cloud computing and software services: theory and techniques**. [S.l.]: CRC Press, 2010. Cited on page 22.

ALBUQUERQUE, J. G. Modelagem e avaliação de desempenho operacional e ambiental em cadeias de suprimentos verdes. In: UFPE. [S.l.], 2013. Cited on page 92.

ALJOUMAH, E. et al. Sla in cloud computing architectures: A comprehensive study. **International Journal of Grid and Distributed Computing**, v. 8, n. 5, p. 7–32, 2015. Cited 2 times on pages 51 and 52.

ALKHALIL, A.; SAHANDI, R.; JOHN, D. A review of the current level of support to aid decisions for migrating to cloud computing. In: **Proceedings of the International Conference on Internet of Things and Cloud Computing**. New York, NY, USA: ACM, 2016. (ICC '16), p. 58:1–58:8. ISBN 978-1-4503-4063-2. Cited 2 times on pages 76 and 83.

ALPENDRE, M. A. C. Service level agreement: A concept to know how to use. In: . [S.l.: s.n.], 2006. Cited 2 times on pages 50 and 51.

ALSHAMAILA, Y.; PAPAGIANNIDIS, S.; LI, F. Cloud computing adoption by smes in the north east of england: A multi-perspective framework. **Journal of Enterprise Information Management**, Emerald Group Publishing Limited, v. 26, n. 3, p. 250–275, 2013. Cited on page 22.

AUSTREGÉSILO, M. S. S.; CALLOU, G. Stochastic models for optimizing availability, cost and sustainability of data center power architectures through genetic algorithm. **Revista de Informática Teórica e Aplicada**, v. 26, n. 2, p. 27–44, 2019. Cited 2 times on pages 77 and 90.

BAGHERI, Z.; ZAMANIFAR, K. Enhancing energy efficiency in resource allocation for real-time cloud services. In: **Telecommunications (IST), 2014 7th International Symposium on**. [S.l.: s.n.], 2014. p. 701–706. Cited on page 75.

BALBO, G. Introduction to stochastic petri nets. In: BRINKSMA, E.; HERMANNS, H.; KATOEN, J.-P. (Ed.). **Lectures on Formal Methods and PerformanceAnalysis**. [S.l.]: Springer Berlin / Heidelberg, 2001, (Lecture Notes in Computer Science, v. 2090). p. 84–155. ISBN 978-3-540-42479-6. Cited on page 48.

BANERJEE, P. et al. Towards a net-zero data center. **ACM Journal on Emerging Technologies in Computing Systems (JETC)**, ACM New York, NY, USA, v. 8, n. 4, p. 1–39, 2012. Cited 3 times on pages 77, 91, and 96.

BELOGLAZOV, A.; ABAWAJY, J.; BUYYA, R. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. **Future Generation Computer Systems**, v. 28, n. 5, p. 755–768, 2012. Cited By 753. Cited on page 75.

BENNACEUR, W. M.; KLOUL, L. Formal models for safety and performance analysis of a data center system. **Reliability Engineering & System Safety**, Elsevier, v. 193, p. 106643, 2020. Cited 3 times on pages 77, 92, and 96.

BENNANI, N. et al. Towards a secure database integration using sla in a multi-cloud context. In: **2015 IEEE 39th Annual Computer Software and Applications Conference**. [S.l.: s.n.], 2015. v. 3, p. 4–9. Cited on page 51.

BERRAL, J. L. et al. Building green cloud services at low cost. In: **Distributed Computing Systems (ICDCS), 2014 IEEE 34th International Conference on**. [S.l.: s.n.], 2014. p. 449–460. ISSN 1063-6927. Cited on page 75.

BLAKE, J. T.; REIBMAN, A. L.; TRIVEDI, K. S. Sensitivity analysis of reliability and performability measures for multiprocessor systems. In: **Proceedings of the 1988 ACM SIGMETRICS conference on Measurement and modeling of computer systems**. [S.l.: s.n.], 1988. p. 177–186. Cited on page 46.

BLISCHKE, W. R.; MURTHY, D. P. **Case studies in reliability and maintenance**. [S.l.]: Wiley Online Library, 2003. v. 480. Cited on page 50.

BOILLAT, T.; LEGNER, C. Why do companies migrate towards cloud enterprise systems? a post-implementation perspective. In: **2014 IEEE 16th Conference on Business Informatics**. [S.l.: s.n.], 2014. v. 1, p. 102–109. ISSN 2378-1963. Cited on page 75.

CALLOU, G.; ANDRADE, E.; FERREIRA, J. Modeling and analyzing availability, cost and sustainability of it data center systems. In: IEEE. **2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)**. [S.l.], 2019. p. 2127–2132. Cited on page 90.

CALLOU, G. et al. An integrated modeling approach to evaluate and optimize data center sustainability, dependability and cost. **Energies**, Multidisciplinary Digital Publishing Institute, v. 7, n. 1, p. 238–277, 2014. Cited 3 times on pages 76, 90, and 102.

CALLOU, G. et al. Sustainability and dependability evaluation on data center architectures. In: IEEE. **2011 IEEE International Conference on Systems, Man, and Cybernetics**. [S.l.], 2011. p. 398–403. Cited 2 times on pages 77 and 90.

CALLOU, G. et al. Estimating sustainability impact of high dependable data centers: a comparative study between brazilian and us energy mixes. **Computing**, Springer, v. 95, n. 12, p. 1137–1170, 2013. Cited 4 times on pages 41, 43, 90, and 122.

CALLOU, G. et al. Estimating sustainability impact of high dependable data centers: a comparative study between brazilian and us energy mixes. **Computing**, Springer, v. 95, n. 12, p. 1137–1170, 2013. Cited 2 times on pages 76 and 90.

CALLOU, G. et al. Impact analysis of maintenance policies on data center power infrastructure. In: IEEE. **2010 IEEE international conference on systems, man and cybernetics**. [S.l.], 2010. p. 526–533. Cited on page 77.

CALLOU, G. et al. A formal approach to the quantification of sustainability and dependability metrics on data center infrastructures. In: **Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium**. [S.l.: s.n.], 2011. p. 274–281. Cited on page 76.

CALLOU, G. R. de A. Assessment to support the planning of sustainable data centers with high availability. In: . [S.l.: s.n.], 2013. Available on: http://www.modcs.org/wp-content/uploads/thesis/Thesis-Gustavo.pdf. Accessed in: July 14, 2020. Cited 3 times on pages 90, 96, and 102.

CAMBOIM, K.; ALENCAR, F. M. Requisitos não funcionais e sustentabilidade para computação em nuvem: uma revisão sistemática da literatura. In: **WER**. [S.l.: s.n.], 2018. Cited on page 139.

CAMBOIM, K. et al. Sustainability analysis in data center dense architectures. 2020. Cited 2 times on pages 122 and 138.

CAMBOIM, K. et al. Availability evaluation and maintenance policy of data center infrastructure. In: SBC. **Anais Estendidos do X Simpósio Brasileiro de Engenharia de Sistemas Computacionais**. [S.l.], 2020. p. 198–203. Cited 4 times on pages 109, 110, 111, and 139.

CARDOSO, J.; VALETTE, R. **Redes de Petri**. Florianópolis: [s.n.], 1997. Cited on page 46.

CASSANDRAS, C.; LAFORTUNE, S. **Introduction to Discrete Event Systems**. [S.l.]: Springer, 2008. Cited on page 42.

CHANG, S. E.; CHIU, K.-M.; CHIAO, Y.-C. Cloud migration: Planning guidelines and execution framework. In: **2015 Seventh International Conference on Ubiquitous and Future Networks**. [S.l.: s.n.], 2015. p. 814–819. ISSN 2165-8528. Cited on page 75.

CHENG, D.; JIANG, C.; ZHOU, X. Heterogeneity-aware workload placement and migration in distributed sustainable datacenters. In: **Parallel and Distributed Processing Symposium, 2014 IEEE 28th International**. [S.l.: s.n.], 2014. p. 307–316. ISSN 1530-2075. Cited on page 75.

CHIOLA, G. et al. Greatspn 1.7: Graphical editor and analyzer for timed and stochastic petri nets. v. 24, p. 47–68, 1995. Cited on page 42.

CHU, W. C. C. et al. An approach of quality of service assurance for enterprise cloud computing (qosaecc). In: **Trustworthy Systems and their Applications (TSA), 2014 International Conference on**. [S.l.: s.n.], 2014. p. 7–13. Cited on page 75.

CHUTE, C. **Regional 2016 SMB Cloud Storage Adoption Survey: Laying the Groundwork for Small Business Disaster Recovery Adoption**. [S.l.]: IDC, 2016. Cited on page 22.

CIARDO, G.; MUPPALA, J.; TRIVEDI, K. **SPNP: stochastic Petri net package**. [S.l.]: IEEE Comput. Soc. Press, 1989. 142–151 p. Cited on page 42.

CISCO. **Cisco 4000 Family Integrated Services Router Data Sheet**. [S.l.], 2020. Available on: https://www.cisco.com/c/en/us/products/collateral/routers/4000-series-integrated-services-routers-isr/data_sheet-c78-732542.html. Accessed in: July 14, 2020. Cited on page 110.

CISCO. **Cisco Catalyst 4500-X Series Fixed 10 Gigabit Ethernet Aggregation Switch Data Sheet**. [S.l.], 2020. Available on: https://www.cisco.com/c/en/us/products/collateral/switches/catalyst-4500-x-series-switches/data_sheet_c78-696791.html. Accessed in: July 14, 2020. Cited on page 110.

DAMASCENO, J. C. Ucloud: Uma abordagem para implantação de nuvem privada para a administração pública federal. In: . [S.l.: s.n.], 2015. Cited on page 137.

DANDRES, T. et al. Consideration of marginal electricity in real-time minimization of distributed data centre emissions. **Journal of Cleaner Production**, v. 143, p. 116 – 124, 2017. ISSN 0959-6526. Cited on page 76.

DÉSOBLIN, G.; PAPINI, H. Sla management: a key differentiator for service providers. 2001. Alcatel Telecommunications Review, 3 rd Quarter. Cited on page 51.

DESROCHERS, A.; AL-JAAR, R. **Applications of Petri Nets in Manufacturing Systems: Modeling, Control, and Performance Analysis**. [S.l.]: IEEE Press, 1995. Cited 2 times on pages 49 and 50.

DIMINICO, C.; JEW, J. **Telecommunications infrastructure standard for data centers ANSI/TIA-942**. 2006. Cited 3 times on pages 32, 33, and 57.

DINCER, I.; ROSEN, M. A. **Exergy: energy, environment and sustainable development**. [S.l.]: Newnes, 2012. Cited on page 41.

EBELING, C. E. **An Introduction to Reliability and Maintainability Engineering**. [S.l.]: University of Dayton, 1997. Cited on page 114.

EROL-KANTARCI, M.; MOUFTAH, H. T. Overlay energy circle formation for cloud data centers with renewable energy futures contracts. In: **2014 IEEE Symposium on Computers and Communications (ISCC)**. [S.l.: s.n.], 2014. Workshops, p. 1–6. ISSN 1530-1346. Cited on page 75.

EXTREME. **Mean Time Between Failures (MTBF)**. [S.l.], 2020. Available on: https://www.extremenetworks.com/support/mean-time-between-failures/Accessed in: July 14, 2020. Cited on page 110.

FERREIRA, J. et al. An algorithm to optimize electrical flows. In: IEEE. **2013 IEEE International Conference on Systems, Man, and Cybernetics**. [S.l.], 2013. p. 109–114. Cited 2 times on pages 76 and 89.

FERREIRA, J. et al. Estimating the environmental impact of data centers. In: IEEE. **2018 IEEE 17th International Symposium on Network Computing and Applications (NCA)**. [S.l.], 2018. p. 1–4. Cited 5 times on pages 41, 42, 76, 89, and 122.

FERREIRA, J. et al. An artificial neural network approach to forecast the environmental impact of data centers. **Information**, Multidisciplinary Digital Publishing Institute, v. 10, n. 3, p. 113, 2019. Cited 2 times on pages 76 and 89.

FERREIRA, J. et al. An algorithm to optimise the energy distribution of data centre electrical infrastructures. **International Journal of Grid and Utility Computing**, Inderscience Publishers (IEL), v. 11, n. 3, p. 419–433, 2020. Cited 4 times on pages 43, 76, 89, and 102.

FERREIRA, J. et al. Pldad—an algorihm to reduce data center energy consumption. **Energies**, Multidisciplinary Digital Publishing Institute, v. 11, n. 10, p. 2821, 2018. Cited 2 times on pages 77 and 89.

FERREIRA, J. et al. An algorithm to optimize electrical flows of private cloud infrastructures. In: IEEE. **2015 IEEE International Conference on Systems, Man, and Cybernetics**. [S.l.], 2015. p. 771–776. Cited 2 times on pages 76 and 89.

FIGUEIRÊDO, J. et al. Estimating reliability importance and total cost of acquisition for data center power infrastructures. In: IEEE. **2011 IEEE International Conference on Systems, Man, and Cybernetics**. [S.l.], 2011. p. 421–426. Cited 2 times on pages 76 and 91.

FIGUEIRÊDO, J. J. C. de. Análise de dependabilidade de sistemas data center baseada em Índices de importância. In: . [S.l.: s.n.], 2011. Available on: http://www.modcs.org/wp-content/uploads/2008/05/jjcf-dissertacao de mestrado.pdf. Accessed in: June 20, 2020. Cited 2 times on pages 91 and 96.

FLORIN, G.; FRAIZE, C.; NATKIN, S. Stochastic petri nets: Properties, applications and tools. **Microelectronics Reliability**, Elsevier, v. 31, n. 4, p. 669–697, 1991. Cited on page 46.

FRANK, P. M.; ESLAMI, M. Introduction to system sensitivity theory. **IEEE Transactions on Systems, Man, and Cybernetics**, IEEE, v. 10, n. 6, p. 337–338, 1980. Cited on page 46.

GENG, H. **Data center handbook**. [S.l.]: John Wiley & Sons, 2014. Cited 9 times on pages 26, 29, 31, 32, 36, 37, 38, 97, and 127.

GERHARDT, D. T. S. T. E. **Métodos de pesquisa**. [S.l.: s.n.], 2009. ISBN 978-85-386-0071-8. Cited 2 times on pages 59 and 122.

GERMAN, R. et al. Timenet - a toolkit for evaluating non-markovian stochastic petri nets. **Performance Evaluation**, v. 24, p. 69–87, 1995. Cited on page 42.

GERMAN, R. et al. Timenet: a toolkit for evaluating non-markovian stochastic petri nets. **Performance Evaluation**, v. 24, p. 69 – 87, 1995. ISSN 0166-5316. Cited on page 47.

GIACOBBE, M. et al. Evaluating a cloud federation ecosystem to reduce carbon footprint by moving computational resources. In: **2015 IEEE Symposium on Computers and Communication (ISCC)**. [S.l.: s.n.], 2015. p. 99–104. Cited on page 75.

GODFREY, L.; TODD, C. Defining thresholds for freshwater sustainability indicators within the context of south african water resource management. 2nd warfa/waternet symposium: Integrated water resource management: Theory, practice, cases. cape town, south africa. In: **2nd WARFA/Waternet Symposium: Integrated Water Resource Management: Theory, Practice, Cases. Cape Town, South Africa**. [S.l.: s.n.], 2001. Cited on page 53.

HAMBY, D. M. A review of techniques for parameter sensitivity analysis of environmental models. **Environmental monitoring and assessment**, Springer, v. 32, n. 2, p. 135–154, 1994. Cited 2 times on pages 45 and 46.

HART, S. L.; MILSTEIN, M. B. Criando valor sustentável. **RAE executivo**, v. 3, n. 2, p. 65–79, 2004. Cited on page 39.

HEISING, C. et al. Ieee recommended practice for the design of reliable industrial and commercial power systems. **IEEE Inc., New York**, 2007. Cited 9 times on pages 29, 35, 36, 37, 38, 118, 127, 134, and 135.

HWANG, J. et al. Enterprise-scale cloud migration orchestrator. In: **2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)**. [S.l.: s.n.], 2015. p. 1002–1007. ISSN 1573-0077. Cited on page 75.

ISLAM, S. et al. Assurance of security and privacy requirements for cloud deployment model. **IEEE Transactions on Cloud Computing**, PP, n. 99, p. 1–1, 2015. ISSN 2168-7161. Cited on page 75.

IV, W. P. T. et al. Tier classification define site infrastructure performance. **Uptime Institute**, v. 17, 2006. Cited 5 times on pages 29, 30, 34, 107, and 119.

JAMES, P. **Urban sustainability in theory and practice: circles of sustainability**. [S.l.]: Routledge, 2014. Cited on page 39.

JANSEN, W.; GRANCE, T. Sp 800-144. guidelines on security and privacy in public cloud computing. National Institute of Standards & Technology, 2011. Cited on page 21.

JÚNIOR, J. F. da S. Estratégias para analisar e estimar a eficiência do consumo de energia em centros de dados. In: . [S.l.: s.n.], 2019. Available on: http://www.modcs.org/wp-content/uploads/thesis/Tese-Joao.pdf. Accessed in: July 14, 2020. Cited 2 times on pages 89 and 96.

JÚNIOR, R. d. S. M. et al. Sensitivity analysis of availability of redundancy in computer networks. **CTRQ 2011**, Citeseer, p. 122, 2011. Cited on page 93.

JÚNIOR, R. de S. M. Identification of availability and performance bottlenecks in cloud computing systems: An approach based on hierarchical models and sensitivity analysis. In: . [S.l.: s.n.], 2016. Available on: http://www.modcs.org/wp-content/uploads/thesis/Tese-RubensMatosJunior.pdf. Accessed in: August 18, 2020. Cited 2 times on pages 93 and 96.

KANDUKURI, B. R.; PATURI, V. R.; RAKSHIT, A. Cloud security issues. In: **Proceedings of the 2009 IEEE International Conference on Services Computing**. Washington, DC, USA: IEEE Computer Society, 2009. (SCC '09), p. 517–520. ISBN 978-0-7695-3811-2. Cited on page 50.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004. Cited 3 times on pages 66, 67, and 68.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. In: KEELE UNIVERSITY AND DURHAM UNIVERSITY JOINT REPORT. **Technical report, Ver. 2.3 EBSE Technical Report. EBSE**. [S.l.]: sn, 2007. Cited 4 times on pages 66, 68, 71, and 139.

KOOMEY, J. Growth in data center electricity use 2005 to 2010. **A report by Analytical Press, completed at the request of The New York Times**, v. 9, 2011. Cited on page 28.

KUO, W.; ZUO, M. J. **Optimal reliability modeling: principles and applications**. [S.l.]: John Wiley & Sons, 2003. Cited 6 times on pages 35, 36, 37, 38, 44, and 128.

LABORATORY, G. M. Annual greenhouse gas index (aggi). In: . [S.l.: s.n.], 2020. Available on: https://www.esrl.noaa.gov/gmd/aggi/aggi.html. Accessed in: September 07, 2020. Cited on page 20.

LEHFELD, N. A. d. S.; BARROS, A. Projeto de pesquisa: propostas metodológicas. **Petrópolis/RJ: Vozes**, 1991. Cited on page 59.

LEIGH, K.; RANGANATHAN, P. Blades as a general-purpose infrastructure for future system architectures: Challenges and solutions. **HP Labs Tech. Rep. HPL-2006-182**, 2007. Cited 2 times on pages 35 and 137.

LI, H.; CASALE, G.; ELLAHI, T. Sla-driven planning and optimization of enterprise applications. In: **Proceedings of the first joint WOSP/SIPEW international conference on Performance engineering**. New York, NY, USA: ACM, 2010. (WOSP/SIPEW '10), p. 117–128. ISBN 978-1-60558-563-5. Cited on page 51.

LILJA, D. J. **Measuring Computer Performance: A Practitioner's Guide**. [S.l.]: Cambridge University Press, 2000. Cited on page 42.

LIU, Y. et al. A colored generalized stochastic petri net simulation model for service reliability evaluation of active-active cloud data center based on it infrastructure. In: IEEE. **2017 2nd International Conference on System Reliability and Safety (ICSRS)**. [S.l.], 2017. p. 51–56. Cited 3 times on pages 76, 94, and 96.

MACIEL, P.; LINS, R.; CUNHA, P. **Introduction of the Petri Net and Applied**. Campinas, SP: [s.n.], 1996. Cited on page 38.

MACIEL, P. et al. Mercury: Performance and dependability evaluation of systems with exponential, expolynomial, and general distributions. In: IEEE. **2017 IEEE 22nd Pacific Rim international symposium on dependable computing (PRDC)**. [S.l.], 2017. p. 50–57. Cited 11 times on pages 42, 43, 44, 64, 101, 118, 121, 128, 133, 157, and 160.

MACIEL, P.; TRIVEDI, K.; KIM, D. Dependability modeling in: Performance and dependability in service computing: Concepts, techniques and research directions. **Hershey: IGI Global, Pennsylvania, USA**, v. 13, 2010. Cited 2 times on pages 35 and 38.

MAGALHÃES, I. L.; PINHEIRO, W. B. **Gerenciamento de serviços de TI na prática: uma abordagem com base na ITIL: inclui ISO/IEC 20.000 e IT Flex**. [S.l.]: Novatec Editora, 2007. Cited 3 times on pages 53, 54, and 56.

MAGOULÈS, F. **Fundamentals of grid computing: theory, algorithms and technologies**. [S.l.]: CRC Press, 2009. Cited on page 28.

MAKOWER, J. A economia verde: descubra as oportunidades e os desafios de uma nova era dos negócios. **São Paulo: Gente**, 2009. Cited on page 40.

MARESOVA, P.; SOBESLAV, V.; KREJCAR, O. Cost–benefit analysis–evaluation model of cloud computing deployment for use in companies. **Applied Economics**, Taylor & Francis, v. 49, n. 6, p. 521–533, 2017. Cited on page 76.

MARIN, P. S. Data centers-desvendando cada passo: conceitos, projeto, infraestrutura física e eficiência energética. **São Paulo: Érica**, 2011. Cited 10 times on pages 23, 29, 30, 31, 35, 97, 98, 99, 100, and 101.

MARSAN, M. A. et al. **Performance models of multiprocessor systems**. [S.l.]: the MIT Press, 1986. Cited 2 times on pages 46 and 49.

MARSAN, M. A. et al. Modelling with generalized stochastic petri nets. J. Wiley & Sons Ltd, 1995. Cited 2 times on pages 48 and 49.

MARWAH, M. et al. Quantifying the sustainability impact of data center availability. **ACM SIGMETRICS Performance Evaluation Review**, ACM New York, NY, USA, v. 37, n. 4, p. 64–68, 2010. Cited 3 times on pages 77, 91, and 96.

MARWAH, M. et al. Data analysis, visualization and knowledge discovery in sustainable data centers. In: **Proceedings of the 2Nd Bangalore Annual Compute Conference**. New York, NY, USA: ACM, 2009. (COMPUTE '09), p. 2:1–2:8. ISBN 978-1-60558-476-8. Cited on page 75.

MATOS, R. et al. Sensitivity analysis of a hierarchical model of mobile cloud computing. **Simulation Modelling Practice and Theory**, Elsevier, v. 50, p. 151–164, 2015. Cited on page 46.

MCCONNELL, J.; SIEGEL, E. **Practical service level management: delivering high-quality web-based services**. [S.l.]: Cisco Press, 2004. Cited on page 50.

MELL, P.; GRANCE, T. et al. The nist definition of cloud computing. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg, 2011. Cited on page 28.

MELO, F. F. L.; JUNIOR, J. F. S.; CALLOU, G. R. de A. Evaluating the impact of maintenance policies associated to sla contracts on the dependability of data centers electrical infrastructures. **Revista de Informática Teórica e Aplicada**, v. 27, n. 1, p. 13–25, 2020. Cited 2 times on pages 76 and 90.

MELO, R. AnÁlise de sensibilidade aplicada À identificaÇÃo de pontos que requerem melhoria na disponibilidade em infraestrura de cloud. In: IEEE. [S.l.], 2017. Cited 2 times on pages 93 and 96.

MOHAMED, I. h.; KARIM, A.; AHMED, A. The migration of the university it infrastructure toward a secure iaas cloud. In: **Electrical and Information Technologies (ICEIT), 2015 International Conference on**. [S.l.: s.n.], 2015. p. 357–362. Cited on page 75.

MOLLOY, M. Performance analysis using stochastic petri nets. **Computers, IEEE Transactions on**, C-31, n. 9, p. 913 –917, sept. 1982. ISSN 0018-9340. Cited on page 47.

MÜLLER, U.; STRUNZ, K. Integrated reliability modeling for data center infrastructures: A case study. In: IEEE. **2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)**. [S.l.], 2012. p. 1–7. Cited 3 times on pages 77, 92, and 96.

MURATA, T. Petri nets: Properties, analysis and applications. **Proceedings of the IEEE**, IEEE, v. 77, n. 4, p. 541–580, 1989. Cited on page 46.

MURUGESAN, S. Harnessing green it: Principles and practices. **IT professional**, IEEE, v. 10, n. 1, 2008. Cited 2 times on pages 22 and 40.

NAKKARIKE NAVEEN ANDA POSEY, M. **U.S. Colocation Services Forecast, 2016–2020**. [S.l.]: IDC, 2016. Available on: http://www.idc.com/getdoc.jsp?containerId=US41284116. Accessed in: October 20, 2020. Cited on page 22.

NASSIF, A. T.; MARTINS, S. A. J. Convergência das redes de comunicação: Aspectos técnicos e econômicos. **Rev. Fac. Ing. - Univ. Tarapacá [online]**, vol.13, n.2, pp. 13-19, 2005. ISSN 0718-1337. Cited on page 51.

NEGLIA, G.; SERENO, M.; BIANCHI, G. Geographical load balancing across green datacenters: A mean field analysis. **ACM SIGMETRICS Performance Evaluation Review**, ACM New York, NY, USA, v. 44, n. 2, p. 64–69, 2016. Cited on page 77.

NGUYEN, T. A. et al. Reliability and availability evaluation for cloud data center networks using hierarchical models. **IEEE Access**, IEEE, v. 7, p. 9273–9313, 2019. Cited 3 times on pages 77, 94, and 96.

NKHOMA, M. Z.; DANG, D. P.; SOUZA-DAW, A. D. Contributing factors of cloud computing adoption: a technology-organisation-environment framework approach. In: **Proceedings of the European Conference on Information Management & Evaluation**. [S.l.: s.n.], 2013. p. 180–189. Cited on page 21.

NUSSBAUMER, N.; LIU, X. Cloud migration for smes in a service oriented approach. In: **Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE 37th Annual**. [S.l.: s.n.], 2013. p. 457–462. Cited 2 times on pages 75 and 94.

PALOS-SÁNCHEZ, P. R.; ARENAS-MÁRQUEZ, F. J.; AGUAYO-CAMACHO, M. Determinants of adoption of cloud computing services by small, medium and large companies. **Journal of Theoretical & Applied Information Technology**, v. 95, n. 6, 2017. Cited on page 76.

PARKHILL, D. F. **The Challenge of the Computer Utility**. [S.l.]: Addison-Wesley Publishing Company, 1966. ISBN 0240507177. Cited on page 28.

PAWLISH, M.; VARDE, A. S.; ROBILA, S. Decision support in data centers for sustainability. In: **2013 IEEE 13th International Conference on Data Mining Workshops**. [S.l.: s.n.], 2013. p. 613–620. ISSN 2375-9232. Cited on page 75.

POPOVIĆ, K.; HOCENSKI, Ž. Cloud computing security issues and challenges. In: IEEE. **MIPRO, 2010 proceedings of the 33rd international convention**. [S.l.], 2010. p. 344–349. Cited on page 21.

PRIYA, B.; PILLI, E. S.; JOSHI, R. C. A survey on energy and power consumption models for greener cloud. In: **Advance Computing Conference (IACC), 2013 IEEE 3rd International**. [S.l.: s.n.], 2013. p. 76–82. Cited on page 75.

QUASCHNING, V. Development of global carbon dioxide emissions and concentration in atmosphere. In: . [S.l.: s.n.], 2020. Available on: https://www.volker-quaschning.de/datserv/CO2/index_e.php. Accessed in: July 24, 2020. Cited on page 20.

QUASCHNING, V. Specific carbon dioxide emissions of various fuels. In: . [S.l.: s.n.], 2020. Available on: https://www.volker-quaschning.de/datserv/CO2-spez/index_e.php. Accessed in: July 24, 2020. Cited 2 times on pages 122 and 124.

RAUSAND, M.; HøYLAND, A. **System reliability theory: models, statistical methods, and applications**. [S.l.: s.n.], 2004. 636 p. Cited 2 times on pages 38 and 44.

RAZA, K.; TURNER, M.; ASAD, S. **CCIE Professional Development: Large Scale IP Network Solutions**. [S.l.]: Cisco Press, 2000. Cited on page 33.

REIS, I. W. Cinvestigação de aspectos verdes na implantação de um data center na Área industrial de suape-pe. In: UNIVERSIDADE FEDERAL DE PERNAMBUCO. [S.l.], 2009. Cited on page 92.

RELIASOFT. Blocksim: System reliability and maintainability analysis software tool. In: . [S.l.: s.n.], 2012. Available on: http://blocksim.reliasoft.com/. Accessed in: October 14, 2020. Cited 2 times on pages 42 and 44.

REN, S. Optimizing water efficiency in distributed data centers. In: **Cloud and Green Computing (CGC), 2013 Third International Conference on**. [S.l.: s.n.], 2013. p. 68–75. Cited on page 75.

RICHTER, R. M. **TI Verde: Sustentabilidade por meio da Computação em Nuvem**. 2013. Cited on page 40.

RITCHIE, A. J.; BROUWER, J. Design of fuel cell powered data centers for sufficient reliability and availability. **Journal of Power Sources**, Elsevier, v. 384, p. 196–206, 2018. Cited 3 times on pages 76, 92, and 96.

RITCHIE, H.; ROSER, M. Co2 and greenhouse gas emissions. In: . [S.l.: s.n.], 2020. Available on: https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions. Accessed in: July 24, 2020. Cited 3 times on pages 19, 20, and 21.

RITCHIE, H.; ROSER, M. Co2 emissions. In: . [S.l.: s.n.], 2020. Available on: https://ourworldindata.org/co2-emissions. Accessed in: July 24, 2020. Cited 2 times on pages 20 and 21.

RITCHIE, H.; ROSER, M. Greenhouse gas emissions. In: . [S.l.: s.n.], 2020. Available on: https://ourworldindata.org/greenhouse-gas-emissions. Accessed in: July 24, 2020. Cited on page 19.

ROSENDO, D. et al. A methodology to assess the availability of next-generation data centers. **The Journal of Supercomputing**, Springer, v. 75, n. 10, p. 6361–6385, 2019. Cited 3 times on pages 76, 94, and 96.

SAHNER, R. A.; TRIVEDI, K. S. Reliability modeling using sharpe. **Reliability, IEEE Transactions on**, R-36, n. 2, p. 186 –193, june 1987. ISSN 0018-9529. Cited 2 times on pages 42 and 44.

SANKAR, S.; GAUTHIER, D.; GURUMURTHI, S. Power availability provisioning in large data centers. In: **Proceedings of the 11th ACM Conference on Computing Frontiers**. [S.l.: s.n.], 2014. p. 1–9. Cited 3 times on pages 77, 92, and 96.

SANTOS, D. L.; CAMBOIM, K.; ALENCAR, F. Avaliação de desempenho de nuvens privadas: Um comparativo entre owncloud, nextcloud e pydio. In: SBC. **Anais do XVI Workshop em Clouds e Aplicações**. [S.l.], 2018. Cited on page 139.

SANTOS, E. J. P. **Método científico: uma introdução – o desafio de ser cientista**. [S.l.: s.n.], 2018. ISBN 978-85-924541-0-4. Cited on page 59.

SCHROEDER, B.; GIBSON, G. A. Understanding disk failure rates: What does an mttf of 1,000,000 hours mean to you? **ACM Transactions on Storage (TOS)**, ACM New York, NY, USA, v. 3, n. 3, p. 8–es, 2007. Cited on page 36.

SCHULZ, G. **The green and virtual data center**. [S.l.]: CRC Press, 2016. Cited 8 times on pages 25, 29, 30, 33, 40, 41, 97, and 98.

SCIENTISTS, U. of C. Why does co2 get most of the attention when there are so many other heat-trapping gases? In: . [S.l.: s.n.], 2020. Available on: https://www.ucsusa.org/resources/why-does-co2-get-more-attention-other-gases. Accessed in: July 15, 2020. Cited on page 20.

SILVAA, S. et al. Dependability evaluation of data center power infrastructures considering substation switching operations. In: **Probabilistic Safety Assessment and Management conference**. [S.l.: s.n.], 2014. Cited 3 times on pages 76, 91, and 96.

SMITH, D. J. **Reliability, maintainability and risk: practical methods for engineers**. [S.l.]: Butterworth-Heinemann, 2017. Cited 4 times on pages 23, 44, 102, and 108.

Smith, W. E. et al. Availability analysis of blade server systems. **IBM Systems Journal**, v. 47, n. 4, p. 621–640, 2008. Cited 6 times on pages 34, 93, 108, 109, 110, and 137.

SONI, M.; NAMJOSHI, J.; PILLAI, S. Robustness and oppurtuneness based approach for cloud deployment model selection. In: **Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on**. [S.l.: s.n.], 2013. p. 207–212. Cited on page 75.

SOUSA, E. et al. A modeling approach for cloud infrastructure planning considering dependability and cost requirements. **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, IEEE, v. 45, n. 4, p. 549–558, 2014. Cited 2 times on pages 76 and 93.

SOUSA, E. et al. Cloud infrastructure planning considering different redundancy mechanisms. **Computing**, Springer, v. 99, n. 9, p. 841–864, 2017. Cited 2 times on pages 76 and 93.

SOUSA, E. T. G. **Avaliação do Impacto de uma Política de Manutenção na Performabilidade de Sistemas de Transferência Eletrônica de Fundos**. Tese (Doutorado) — Universidade Federal de Pernambuco, 2009. Cited on page 50.

SOUSA, E. T. G. d. Modelagem de desempenho, dependabilidade e custo para o planejamento de infraestruturas de nuvens privadas. In: UNIVERSIDADE FEDERAL DE PERNAMBUCO. [S.l.], 2015. Cited 2 times on pages 93 and 96.

SOUZA, R. R. de. Avaliação de dependabilidade de infraestruturas de datacenters considerando os efeitos da variação detemperatura. In: . [S.l.: s.n.], 2013. Available on: http://www.modcs.org/wp-content/uploads/thesis/Dissertacao-Rafael.pdf. Accessed in: August 22, 2020. Cited 2 times on pages 91 and 96.

SRIPANIDKULCHAI, K.; SUJICHANTARARAT, S. A business-driven framework for evaluating cloud computing. In: **2012 IEEE Network Operations and Management Symposium**. [S.l.: s.n.], 2012. p. 1335–1342. ISSN 1542-1201. Cited on page 75.

START. State of the art through systematic review. In: . [S.l.: s.n.], 2020. Available on: http://lapes.dc.ufscar.br/tools/start_tool. Accessed in: November 14, 2020. Cited 4 times on pages 64, 89, 155, and 156.

STAS, F. Available on: Playtime: Hp blade center - part 2: Woops. In: . [S.l.: s.n.], 2020. Https://blog.suddenelfilio.net/playtime-hp-blade-center-part-2-woops/. Accessed in: July 02, 2020. Cited on page 35.

STURM, R.; MORRIS, W.; JANDER, M. **Foundations of Service Level Management**. [S.l.]: Sams Publishing, 2000. Cited on page 50.

SUN, K.; LI, Y. Effort estimation in cloud migration process. In: **Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on**. [S.l.: s.n.], 2013. p. 84–91. Cited on page 75.

SVERDLIK, Y. Study: Data centers responsible for 1 percent of all electricity consumed worldwide. In: **Data Center Knoledge**. [S.l.: s.n.], 2020. Available on: https://www.datacenterknowledge.com/energy/study-data-centers-responsible-1-percent-all-electricity-consumed-worldwide. Accessed in: June 09, 2020. Cited 2 times on pages 22 and 136.

TRIVEDI, K. et al. Dependability and security models. In: **Design of Reliable Communication Networks, 2009. DRCN 2009. 7th International Workshop on**. [S.l.: s.n.], 2009. p. 11 –20. Cited 3 times on pages 35, 37, and 39.

TRIVEDI, K.; MALHOTRA, M. Reliability and performability techniques and tools: A survey. In: **Messung, Modellierung und Bewertung von Rechenund Kommunikation ssystemen**. [S.l.: s.n.], 1993. p. 27 – 48. Cited 2 times on pages 44 and 45.

TSALIS, N.; THEOHARIDOU, M.; GRITZALIS, D. Return on security investment for cloud platforms. In: **2013 IEEE 5th International Conference on Cloud Computing Technology and Science**. [S.l.: s.n.], 2013. v. 2, p. 132–137. Cited on page 75.

VALE, K. M. A. C.; ALENCAR, F. M. R. de. Challenges, patterns and sustainability indicators for cloud computing. **Brazilian Journal of Development**, v. 6, n. 8, p. 57031–57053, 2020. Cited 3 times on pages 22, 53, and 139.

VALENTIM, T.; CALLOU, G. A model-based strategy for quantifying the impact of availability on the energy flow of data centers. **The Journal of Supercomputing**, Springer, p. 1–24, 2020. Cited 2 times on pages 76 and 90.

VALLE, C. E. do. **Qualidade Ambiental-ISO 14.000**. [S.l.]: Senac, 2002. Cited on page 57.

VAQUERO, L. M. et al. A break in the clouds: towards a cloud definition. **ACM SIGCOMM Computer Communication Review**, ACM, v. 39, n. 1, p. 50–55, 2008. Cited on page 28.

VERAS, M. **Datacenter Componente Central da Infraestrutura de TI (Tecnologia da Informação)**. [S.l.]: Rio de Janeiro: Brasport Livros/Multimídia Ltda, 2009. Cited 4 times on pages 29, 32, 33, and 107.

VERAS, M. **Cloud Computing: nova arquitetura da TI**. [S.l.]: Brasport, 2012. Cited on page 28.

VHP. **ISO/IEC 20000: Uma introdução**. [S.l.]: Van Haren Publishing, 2009. Cited 2 times on pages 55 and 56.

WANG, H.; PHAM, H. **Reliability and Optimal Maintenance**. [S.l.]: Springer Verlag, 2006. Cited on page 50.

WANG, L. et al. Cloud computing: a perspective study. **New Generation Computing**, Springer, v. 28, n. 2, p. 137–146, 2010. Cited on page 21.

WCE&D, U. Un world comission on environment and development. **Our Common Future**, Oxford University Press Oxford, 1987. Cited on page 39.

WERNER, J. et al. Simulator improvements to validate the green cloud computing approach. In: **Network Operations and Management Symposium (LANOMS), 2011 7th Latin American**. [S.l.: s.n.], 2011. p. 1–8. Cited 2 times on pages 75 and 83.

WITKOWSKI, M. et al. Enabling sustainable clouds via environmentally opportunistic computing. In: **Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on**. [S.l.: s.n.], 2010. p. 587–592. Cited on page 75.

XIE, M.; POH, K.-L.; DAI, Y.-S. **Computing System Reliability: Models and Analysis**. [S.l.]: Springer, 2004. Cited 3 times on pages 38, 39, and 44.

YAM, C.-Y. et al. Migration to cloud as real option: Investment decision under uncertainty. In: IEEE. **Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on**. [S.l.], 2011. p. 940–949. Cited on page 21.

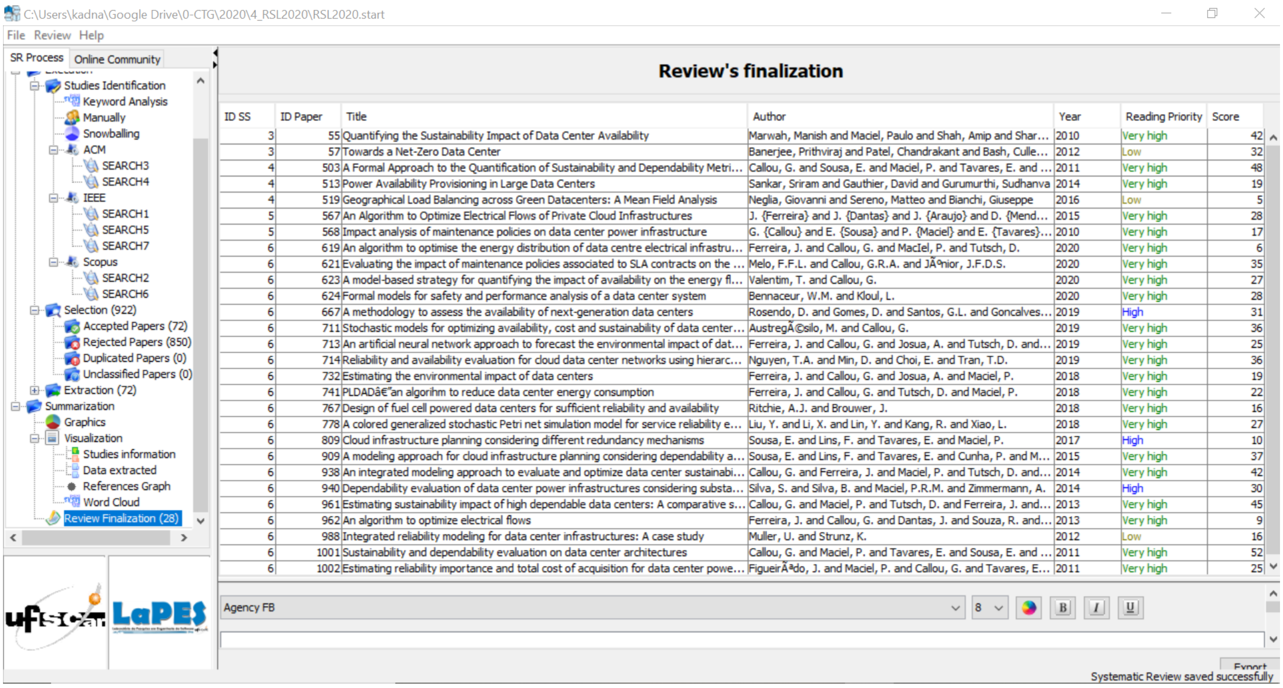
YOUNGE, A. J. et al. Efficient resource management for cloud computing environments. In: **Green Computing Conference, 2010 International**. [S.l.: s.n.], 2010. p. 357–364. Cited on page 92.

YOUNGE, A. J. et al. Efficient resource management for cloud computing environments. In: **Proceedings of the International Conference on Green Computing**. USA: IEEE Computer Society, 2010. (GREENCOMP '10), p. 357–364. ISBN 9781424476121. Cited on page 75.

APPENDIX A – START TOOL

StArt[1] is a software to assist in the conduct of SLR[2], developed by LAPES[3]. It is currently in version BETA 30.3. Figure 40 shows the StArt interface with a total of 28 articles that were classified as related to this research. Figure 41 shows the number of articles per search base. Of the 922 papers, 473 are from ACM, 8 from IEEE, and 441 from Scopus.

Figure 40 – StArt Tool - SLR Finalization for Related Works.

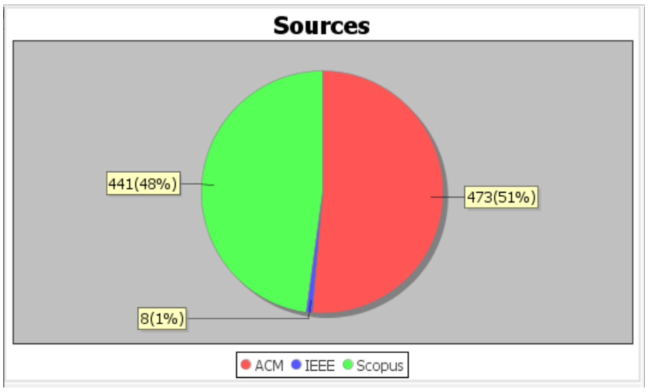


ID SS	ID Paper	Title	Author	Year	Reading Priority	Score
3	55	Quantifying the Sustainability Impact of Data Center Availability	Marwah, Manish and Maciel, Paulo and Shah, Amip and Shar…	2010	Very high	42
3	57	Towards a Net-Zero Data Center	Banerjee, Prithviraj and Patel, Chandrakant and Bash, Culle…	2012	Low	32
4	503	A Formal Approach to the Quantification of Sustainability and Dependability Metri…	Callou, G. and Sousa, E. and Maciel, P. and Tavares, E. and …	2011	Very high	48
4	513	Power Availability Provisioning in Large Data Centers	Sankar, Sriram and Gauthier, David and Gurumurthi, Sudhanva	2014	Very high	19
4	519	Geographical Load Balancing across Green Datacenters: A Mean Field Analysis	Neglia, Giovanni and Sereno, Matteo and Bianchi, Giuseppe	2016	Low	5
5	567	An Algorithm to Optimize Electrical Flows of Private Cloud Infrastructures	J. {Ferreira} and J. {Dantas} and J. {Araujo} and D. {Mend…	2015	Very high	28
5	568	Impact analysis of maintenance policies on data center power infrastructure	G. {Callou} and E. {Sousa} and P. {Maciel} and E. {Tavares}…	2010	Very high	17
6	619	An algorithm to optimise the energy distribution of data centre electrical infrastru…	Ferreira, J. and Callou, G. and MacIel, P. and Tutsch, D.	2020	Very high	6
6	621	Evaluating the impact of maintenance policies associated to SLA contracts on the …	Melo, F.F.L. and Callou, G.R.A. and JÃºnior, J.F.D.S.	2020	Very high	35
6	623	A model-based strategy for quantifying the impact of availability on the energy fl…	Valentim, T. and Callou, G.	2020	Very high	27
6	624	Formal models for safety and performance analysis of a data center system	Bennaceur, W.M. and Kloul, L.	2020	Very high	28
6	667	A methodology to assess the availability of next-generation data centers	Rosendo, D. and Gomes, D. and Santos, G.L. and Goncalves…	2019	High	31
6	711	Stochastic models for optimizing availability, cost and sustainability of data center…	AustregÃ©silo, M. and Callou, G.	2019	Very high	36
6	713	An artificial neural network approach to forecast the environmental impact of dat…	Ferreira, J. and Callou, G. and Josua, A. and Tutsch, D. and…	2019	Very high	25
6	714	Reliability and availability evaluation for cloud data center networks using hierarc…	Nguyen, T.A. and Min, D. and Choi, E. and Tran, T.D.	2019	Very high	36
6	732	Estimating the environmental impact of data centers	Ferreira, J. and Callou, G. and Josua, A. and Maciel, P.	2018	Very high	19
6	741	PLDADâ€"an algorihm to reduce data center energy consumption	Ferreira, J. and Callou, G. and Tutsch, D. and Maciel, P.	2018	Very high	22
6	767	Design of fuel cell powered data centers for sufficient reliability and availability	Ritchie, A.J. and Brouwer, J.	2018	Very high	16
6	778	A colored generalized stochastic Petri net simulation model for service reliability e…	Liu, Y. and Li, X. and Lin, Y. and Kang, R. and Xiao, L.	2018	Very high	27
6	809	Cloud infrastructure planning considering different redundancy mechanisms	Sousa, E. and Lins, F. and Tavares, E. and Maciel, P.	2017	High	10
6	909	A modeling approach for cloud infrastructure planning considering dependability a…	Sousa, E. and Lins, F. and Tavares, E. and Cunha, P. and M…	2015	Very high	37
6	938	An integrated modeling approach to evaluate and optimize data center sustainabi…	Callou, G. and Ferreira, J. and Maciel, P. and Tutsch, D. and…	2014	Very high	42
6	940	Dependability evaluation of data center power infrastructures considering substa…	Silva, S. and Silva, B. and Maciel, P.R.M. and Zimmermann, A.	2014	High	30
6	961	Estimating sustainability impact of high dependable data centers: A comparative s…	Callou, G. and Maciel, P. and Tutsch, D. and Ferreira, J. and…	2013	Very high	45
6	962	An algorithm to optimize electrical flows	Ferreira, J. and Callou, G. and Dantas, J. and Souza, R. and…	2013	Very high	9
6	988	Integrated reliability modeling for data center infrastructures: A case study	Muller, U. and Strunz, K.	2012	Low	16
6	1001	Sustainability and dependability evaluation on data center architectures	Callou, G. and Maciel, P. and Tavares, E. and Sousa, E. and …	2011	Very high	52
6	1002	Estimating reliability importance and total cost of acquisition for data center powe…	FigueirÃªdo, J. and Maciel, P. and Callou, G. and Tavares, E…	2011	Very high	25

Source – StArt Tool (START, 2020).

1 http://lapes.dc.ufscar.br/tools/start_tool
2 Systematic Literature Review
3 Laboratory of Research on Software Engineering

Figure 41 – StArt Tool - Related Works by Search Bases.

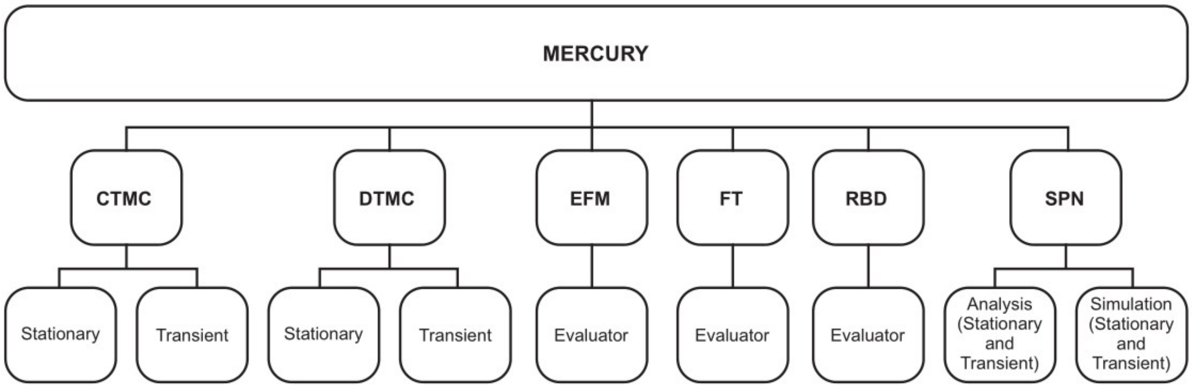


Source – StArt Tool (START, 2020).

APPENDIX B – MERCURY TOOL

Mercury[4] has been developed by MoDCS[5] Research Group at Informatics Center (CIn) of the Federal University of Pernambuco (UFPE) in Brazil since 2009 (MACIEL et al., 2017). An overview of Mercury features is presented in Figure 42.

Figure 42 – Mercury Tool - Features.

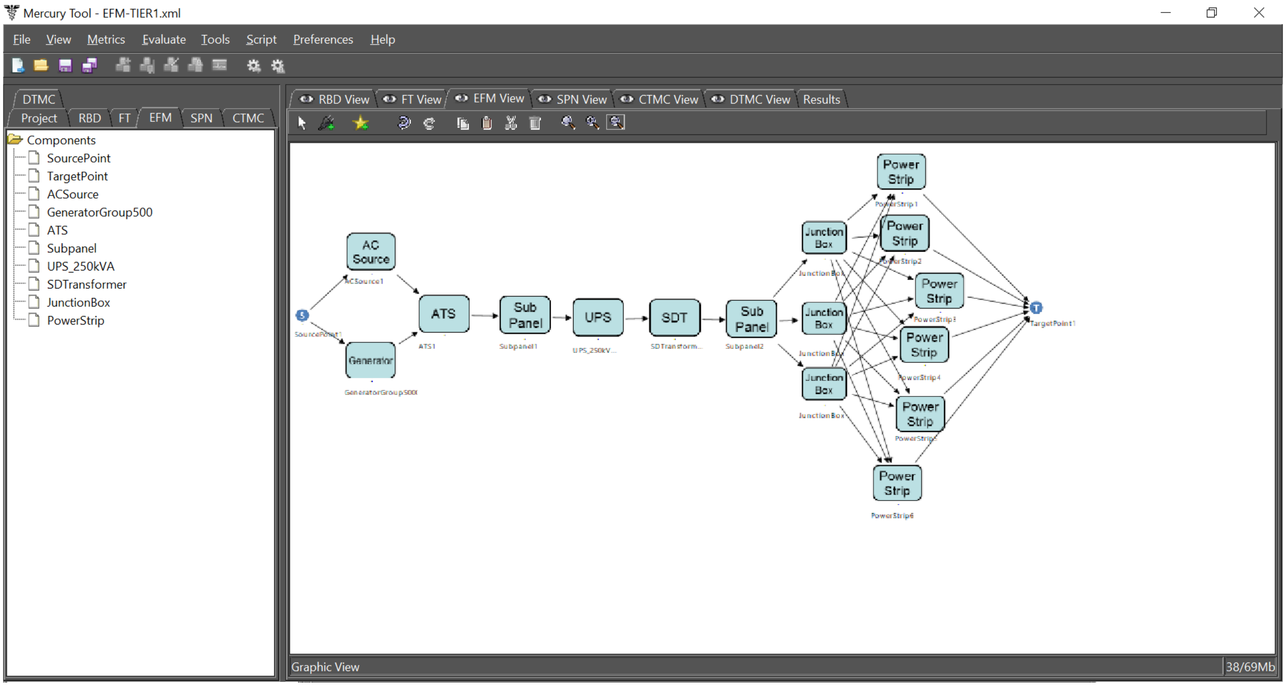


Source – Mercury Tool (MACIEL et al., 2017).

It is worth mentioning that, for simplification (size reduction), all the models presented in the document were designed in another graphic editing tool. Figures 43 to 46 show the EFM models created in the Mercury Tool. Figure 47 presents the results interface of the EFM model in the Mercury Tool. If there are any interconnection errors or component values in the model, the result is given as false. Figure 48 shows the interface for inserting/modifying parameters of an RBD block.

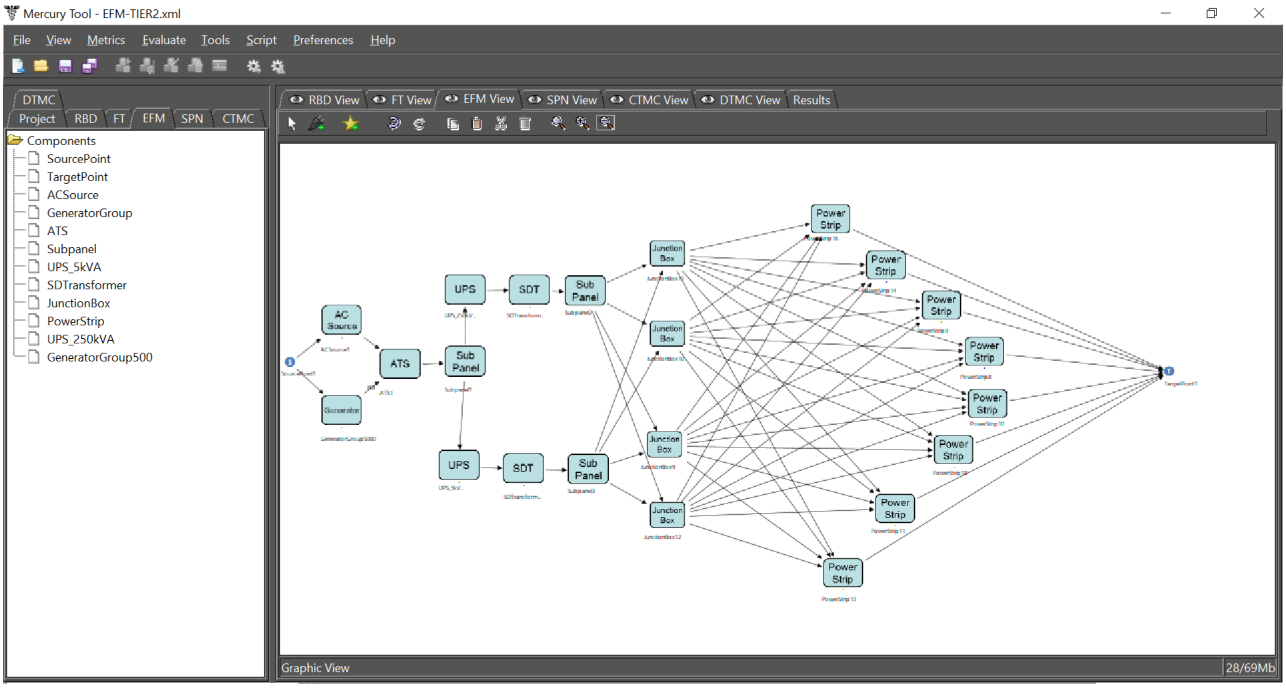
4 http://www.modcs.org/?page_id=2392
5 Modeling of Distributed and Concurrent Systems

Figure 43 – Mercury Tool - EFM Tier 1.



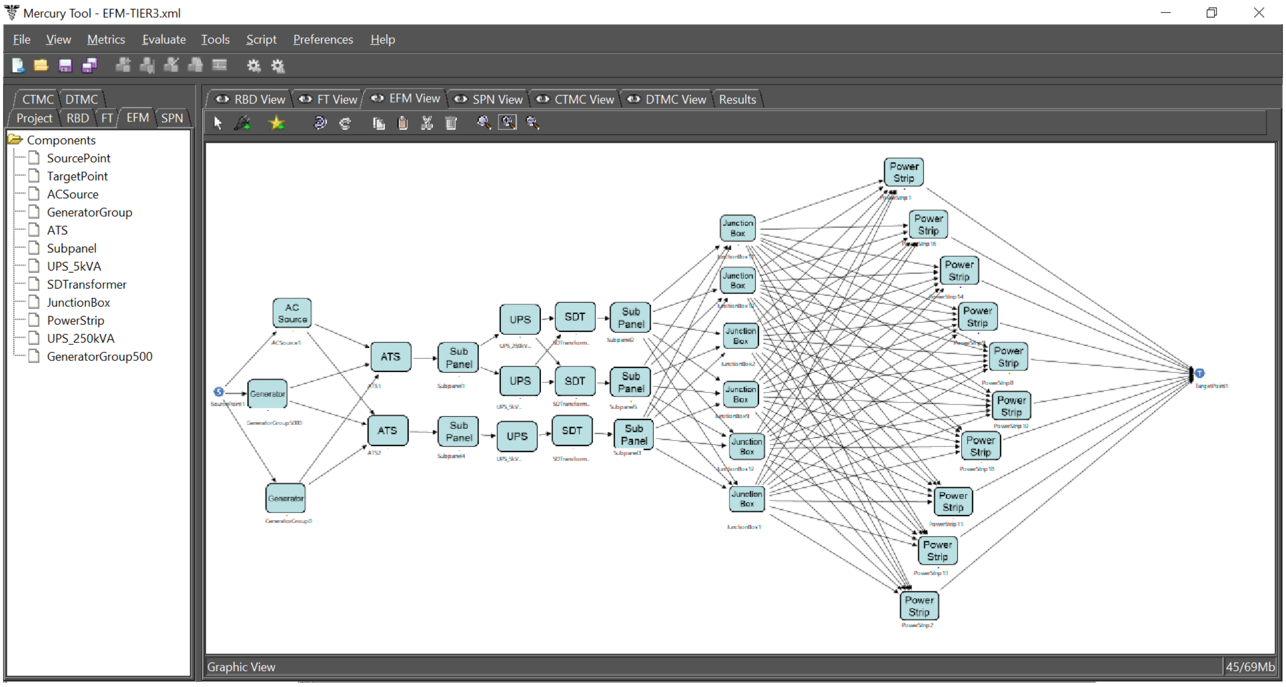
Source – Elaborated by the author from the Mercury Tool.

Figure 44 – Mercury Tool - EFM Tier 2.



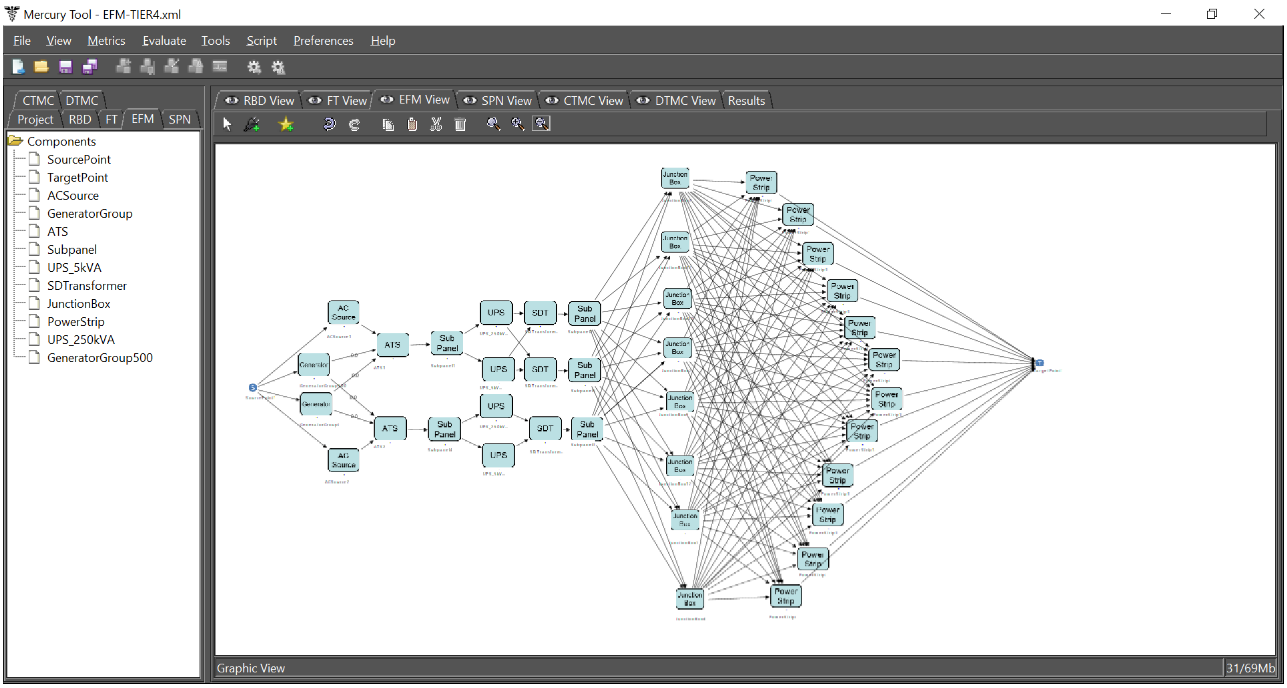
Source – Elaborated by the author from the Mercury Tool.

Figure 45 – Mercury Tool - EFM Tier 3.



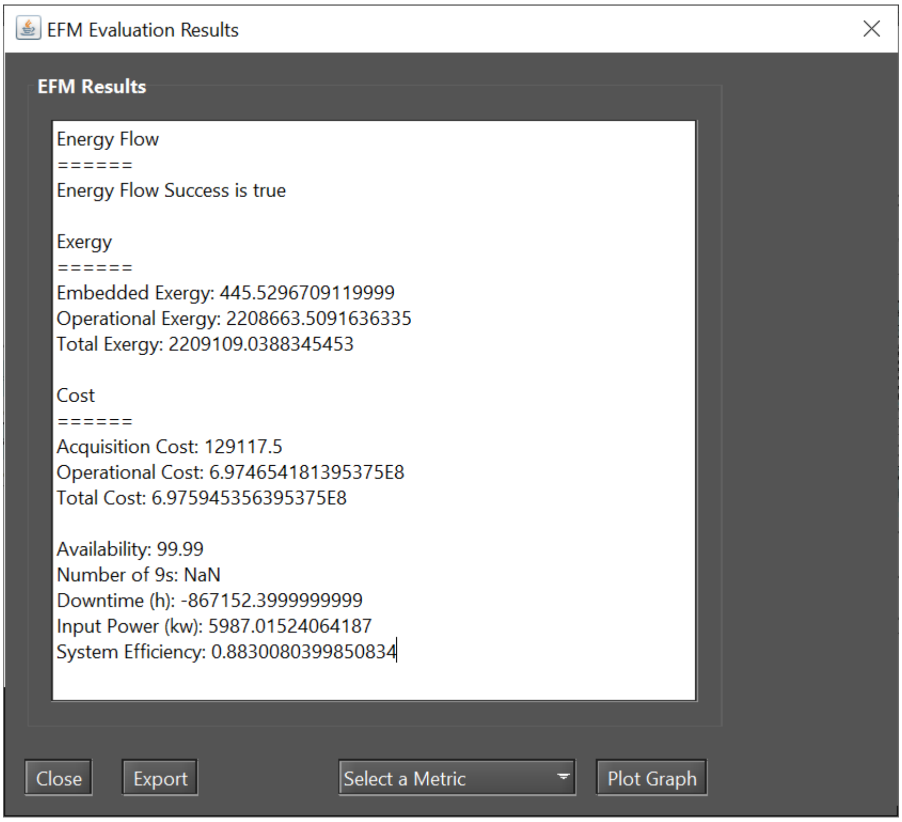
Source – Elaborated by the author from the Mercury Tool.

Figure 46 – Mercury Tool - EFM Tier 4.



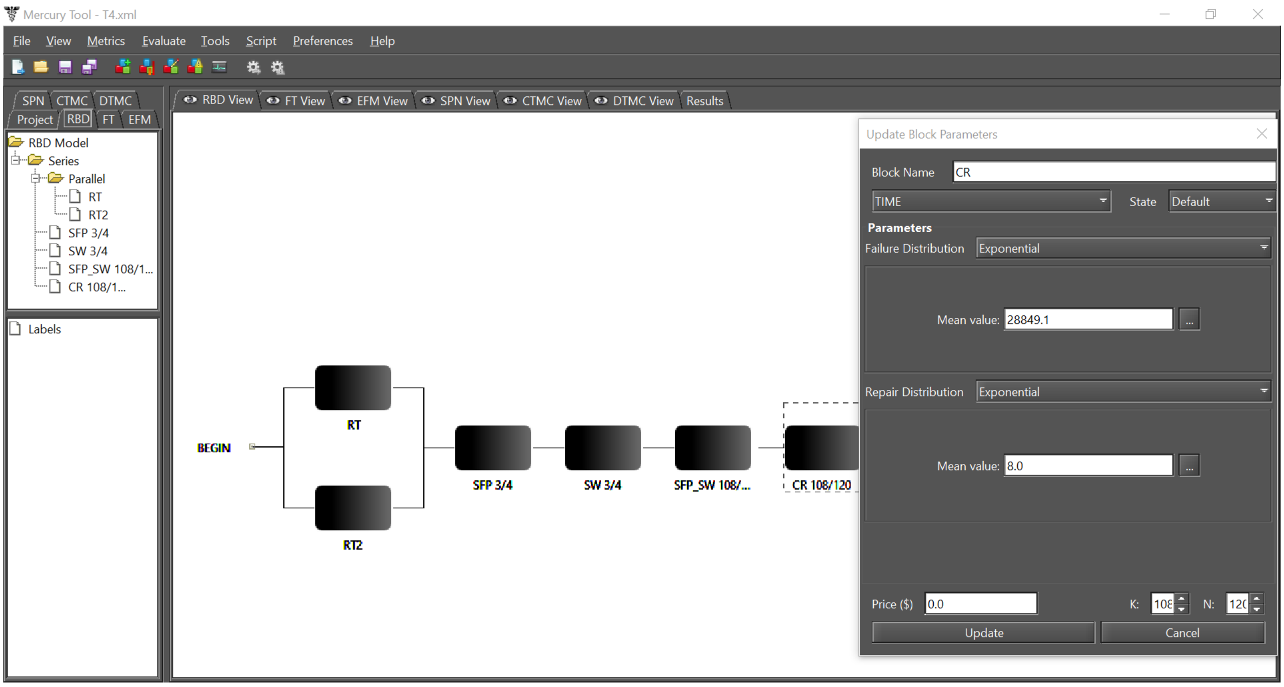
Source – Elaborated by the author from the Mercury Tool.

Figure 47 – Mercury Tool - EFM Result.



Source – Mercury Tool (MACIEL et al., 2017).

Figure 48 – Mercury Tool - Parameters of an RBD Component.



Source – Mercury Tool (MACIEL et al., 2017).

APPENDIX C – SENSITIVITY ANALYSIS SCRIPTS

To perform the parametric sensitivity analysis, the script language supported by the Mercury Tool was used. In this appendix, the scripts developed are presented. Listings from C.1 to C.10 show the scripts in full. Thus, it is possible to observe, as explained in Chapter 5, the parameters used in each analysis and the MTTFs and MTTRs variation for each component.

Listing C.1 – Mercury Script for RBD Sensitivity Analysis for Tier 1 Electrical Architecture.

```
mttf_ACSource = 4380;
mttr_ACSource = 8;
mttf_Generator = 2500;
mttr_Generator = 8;
mttf_ATS = 24038;
mttr_ATS = 8;
mttf_Subpanel1 = 152000;
mttr_Subpanel1 = 8;
mttf_UPS1 = 50000;
mttr_UPS1 = 8;
mttf_SDT1 = 141290;
mttr_SDT1 = 8;
mttf_Subpanel2 = 152000;
mttr_Subpanel2 = 8;
mttf_JB1 = 522400;
mttr_JB1 = 8;
mttf_JB2 = 522400;
mttr_JB2 = 8;
mttf_JB3 = 522400;
mttr_JB3 = 8;
mttf_PS1 = 215111;
mttr_PS1 = 8;
mttf_PS2 = 215111;
mttr_PS2 = 8;
mttf_PS3 = 215111;
mttr_PS3 = 8;
mttf_PS4 = 215111;
mttr_PS4 = 8;
mttf_PS5 = 215111;
mttr_PS5 = 8;
mttf_PS6 = 215111;
mttr_PS6 = 8;
RBD Model{
   block ACSource( MTTF = mttf_ACSource, MTTR = mttr_ACSource);
   block Generator( MTTF = mttf_Generator, MTTR = mttr_Generator);
   parallel p1(ACSource, Generator);
   block ATS( MTTF = mttf_ATS, MTTR = mttr_ATS);
```

```
   block Subpanel1( MTTF = mttf_Subpanel1, MTTR = mttr_Subpanel1);
   block UPS1( MTTF = mttf_UPS1, MTTR = mttr_UPS1);
   block SDT1( MTTF = mttf_SDT1, MTTR = mttr_SDT1);
   block Subpanel2( MTTF = mttf_Subpanel2, MTTR = mttr_Subpanel2);
   block JB1( MTTF = mttf_JB1, MTTR = mttr_JB1);
   block JB2( MTTF = mttf_JB2, MTTR = mttr_JB2);
   block JB3( MTTF = mttf_JB3, MTTR = mttr_JB3);
   parallel p3(JB1, JB2, JB3);
   series s2(UPS1, SDT1, Subpanel2, p3 );
   block PS1( MTTF = mttf_PS1, MTTR = mttr_PS1);
   block PS2( MTTF = mttf_PS2, MTTR = mttr_PS2);
   block PS3( MTTF = mttf_PS3, MTTR = mttr_PS3);
   parallel p4(PS1, PS2, PS3);
   block PS4( MTTF = mttf_PS4, MTTR = mttr_PS4);
   block PS5( MTTF = mttf_PS5, MTTR = mttr_PS5);
   block PS6( MTTF = mttf_PS6, MTTR = mttr_PS6);
   parallel p5(PS4, PS5, PS6);
   series s0(p1, ATS, Subpanel1, s2, p4, p5 );
   top s0;
   metric av = availability;
}  main{ av = solve(Model, av);
  println(av);
  percentageDifference(
            model_ = "Model",
            metric_ = "av",
            samplingPoints = 5,
            parameters =  (
                        mttf_ACSource = [ 2190, 6570 ],
                        mttr_ACSource = [ 4, 12 ],
                        mttf_Generator = [ 1250, 3750 ],
                        mttr_Generator = [ 4, 12 ],
                        mttf_ATS = [ 12019, 36057 ],
                        mttr_ATS = [ 4, 12 ],
                        mttf_Subpanel1 = [ 76000, 228000 ],
                        mttr_Subpanel1 = [ 4, 12 ],
                        mttf_UPS1 = [ 25000, 75000 ],
                        mttr_UPS1 = [ 4, 12 ],
                        mttf_SDT1 = [ 70645, 211935 ],
                        mttr_SDT1 = [ 4, 12 ],
                        mttf_Subpanel2 = [ 76000, 228000 ],
                        mttr_Subpanel2 = [ 4, 12 ],
                        mttf_JB1 = [ 261200, 783600 ],
                        mttr_JB1 = [ 4, 12 ],
                        mttf_JB2 = [ 261200, 783600 ],
                        mttr_JB2 = [ 4, 12 ],
                        mttf_JB3 = [ 261200, 783600 ],
                        mttr_JB3 = [ 4, 12 ],
```

```
                                mttf_PS1 = [ 107555, 322666 ],
                                mttr_PS1 = [ 4, 12 ],
                                mttf_PS2 = [ 107555, 322666 ],
                                mttr_PS2 = [ 4, 12 ],
                                mttf_PS3 = [ 107555, 322666 ],
                                mttr_PS3 = [ 4, 12 ],
                                mttf_PS4 = [ 107555, 322666 ],
                                mttr_PS4 = [ 4, 12 ],
                                mttf_PS5 = [ 107555, 322666 ],
                                mttr_PS5 = [ 4, 12 ],
                                mttf_PS6 = [ 107555, 322666 ],
                                mttr_PS6 = [ 4, 12 ] ),
                output = ( type = "swing",
                           yLabel = "Steady-state availability",
                           baselineValue = av )); }
```

Listing C.2 – Mercury Script for RBD Sensitivity Analysis for Tier 2 Electrical Architecture.

```
mttf_ACSource = 4380;
mttr_ACSource = 8;
mttf_Generator = 2500;
mttr_Generator = 8;
mttf_ATS = 24038;
mttr_ATS = 8;
mttf_Subpanel1 = 152000;
mttr_Subpanel1 = 8;
mttf_UPS1 = 50000;
mttr_UPS1 = 8;
mttf_SDT1 = 141290;
mttr_SDT1 = 8;
mttf_Subpanel2 = 152000;
mttr_Subpanel2 = 8;
mttf_UPS2 = 50000;
mttr_UPS2 = 8;
mttf_SDT2 = 141290;
mttr_SDT2 = 8;
mttf_Subpanel3 = 152000;
mttr_Subpanel3 = 8;
mttf_JB1 = 522400;
mttr_JB1 = 8;
mttf_JB2 = 522400;
mttr_JB2 = 8;
mttf_JB3 = 522400;
mttr_JB3 = 8;
mttf_JB4 = 522400;
mttr_JB4 = 8;
mttf_PS1 = 215111;
mttr_PS1 = 8;
```

```
mttf_PS2 = 215111;
mttr_PS2 = 8;
mttf_PS3 = 215111;
mttr_PS3 = 8;
mttf_PS4 = 215111;
mttr_PS4 = 8;
mttf_PS5 = 215111;
mttr_PS5 = 8;
mttf_PS6 = 215111;
mttr_PS6 = 8;
mttf_PS7 = 215111;
mttr_PS7 = 8;
mttf_PS7 = 215111;
mttr_PS7 = 8;
mttf_PS8 = 215111;
mttr_PS8 = 8;
RBD Model{
   block ACSource( MTTF = mttf_ACSource, MTTR = mttr_ACSource);
   block Generator( MTTF = mttf_Generator, MTTR = mttr_Generator);
   parallel p1(ACSource, Generator);
   block ATS( MTTF = mttf_ATS, MTTR = mttr_ATS);
   block Subpanel1( MTTF = mttf_Subpanel1, MTTR = mttr_Subpanel1);
   block UPS1( MTTF = mttf_UPS1, MTTR = mttr_UPS1);
   block SDT1( MTTF = mttf_SDT1, MTTR = mttr_SDT1);
   block Subpanel2( MTTF = mttf_Subpanel2, MTTR = mttr_Subpanel2);
   series s3(UPS1, SDT1, Subpanel2 );
   block UPS2(  MTTF = mttf_UPS2, MTTR = mttr_UPS2);
   block SDT2(  MTTF = mttf_SDT2, MTTR = mttr_SDT2);
   block Subpanel3( MTTF = mttf_Subpanel3, MTTR = mttr_Subpanel3);
   series s4(UPS2, SDT2, Subpanel3 );
   parallel p2(s3, s4);
   block JB1( MTTF = mttf_JB1, MTTR = mttr_JB1);
   block JB2( MTTF = mttf_JB2, MTTR = mttr_JB2);
   parallel p5(JB1, JB2);
   block JB3( MTTF = mttf_JB3, MTTR = mttr_JB3);
   block JB4( MTTF = mttf_JB4, MTTR = mttr_JB4);
   parallel p6(JB3, JB4);
   block PS1( MTTF = mttf_PS1, MTTR = mttr_PS1);
   block PS2( MTTF = mttf_PS2, MTTR = mttr_PS2);
   block PS3( MTTF = mttf_PS3, MTTR = mttr_PS3);
   parallel p7(PS1, PS2, PS3);
   block PS4( MTTF = mttf_PS4, MTTR = mttr_PS4);
   block PS5( MTTF = mttf_PS5, MTTR = mttr_PS5);
   parallel p8(PS4, PS5);
   block PS6( MTTF = mttf_PS6, MTTR = mttr_PS6);
   block PS7( MTTF = mttf_PS7, MTTR = mttr_PS7);
   block PS8( MTTF = mttf_PS8, MTTR = mttr_PS8);
```

```
    parallel p9(PS6, PS7, PS8);
    series s0(p1, ATS, Subpanel1, p2, p5, p6, p7, p8, p9 );
    top s0;
    metric av = availability;
} main{ av = solve(Model, av);
      println(av);
  percentageDifference(
            model_ = "Model",
            metric_ = "av",
            samplingPoints = 5,
            parameters =  (
                        mttf_ACSource = [ 2190, 6570 ],
                        mttr_ACSource = [ 4, 12 ],
                        mttf_Generator = [ 1250, 3750 ],
                        mttr_Generator = [ 4, 12 ],
                        mttf_ATS = [ 12019, 36057 ],
                        mttr_ATS = [ 4, 12 ],
                        mttf_Subpanel1 = [ 76000, 228000 ],
                        mttr_Subpanel1 = [ 4, 12 ],
                        mttf_UPS1 = [ 25000, 75000 ],
                        mttr_UPS1 = [ 4, 12 ],
                        mttf_SDT1 = [ 70645, 211935 ],
                        mttr_SDT1 = [ 4, 12 ],
                        mttf_Subpanel2 = [ 76000, 228000 ],
                        mttr_Subpanel2 = [ 4, 12 ],
                        mttf_UPS2 = [ 25000, 75000 ],
                        mttr_UPS2 = [ 4, 12 ],
                        mttf_SDT2 = [ 70645, 211935 ],
                        mttr_SDT2 = [ 4, 12 ],
                        mttf_Subpanel3 = [ 76000, 228000 ],
                        mttr_Subpanel3 = [ 4, 12 ],
                        mttf_JB1 = [ 261200, 783600 ],
                        mttr_JB1 = [ 4, 12 ],
                        mttf_JB2 = [ 261200, 783600 ],
                        mttr_JB2 = [ 4, 12 ],
                        mttf_JB3 = [ 261200, 783600 ],
                        mttr_JB3 = [ 4, 12 ],
                        mttf_JB4 = [ 261200, 783600 ],
                        mttr_JB4 = [ 4, 12 ],
                        mttf_PS1 = [ 107555, 322666 ],
                        mttr_PS1 = [ 4, 12 ],
                        mttf_PS2 = [ 107555, 322666 ],
                        mttr_PS2 = [ 4, 12 ],
                        mttf_PS3 = [ 107555, 322666 ],
                        mttr_PS3 = [ 4, 12 ],
                        mttf_PS4 = [ 107555, 322666 ],
                        mttr_PS4 = [ 4, 12 ],
```

```
                        mttf_PS5 = [ 107555, 322666 ],
                        mttr_PS5 = [ 4, 12 ],
                        mttf_PS6 = [ 107555, 322666 ],
                        mttr_PS6 = [ 4, 12 ],
                        mttf_PS7 = [ 107555, 322666 ],
                        mttr_PS7 = [ 4, 12 ],
                        mttf_PS8 = [ 107555, 322666 ],
                        mttr_PS8 = [ 4, 12 ] ),
            output = ( type = "swing",
                       yLabel = "Steady-state availability",
                       baselineValue = av ));}
```

Listing C.3 – Mercury Script for RBD Sensitivity Analysis for Tier 3 Electrical Architecture.

```
mttf_ACSource = 4380;
mttr_ACSource = 8;
mttf_Generator = 2500;
mttr_Generator = 8;
mttf_Generator2 = 2500;
mttr_Generator2 = 8;
mttf_ATS = 24038;
mttr_ATS = 8;
mttf_Subpanel1 = 152000;
mttr_Subpanel1 = 8;
mttf_UPS1 = 50000;
mttr_UPS1 = 8;
mttf_UPS2 = 50000;
mttr_UPS2 = 8;
mttf_SDT1 = 141290;
mttr_SDT1 = 8;
mttf_Subpanel2 = 152000;
mttr_Subpanel2 = 8;
mttf_SDT2 = 141290;
mttr_SDT2 = 8;
mttf_Subpanel3 = 152000;
mttr_Subpanel3 = 8;
mttf_ATS2 = 24038;
mttr_ATS2 = 8;
mttf_Subpanel4 = 152000;
mttr_Subpanel4 = 8;
mttf_UPS3 = 50000;
mttr_UPS3 = 8;
mttf_SDT3 = 141290;
mttr_SDT3 = 8;
mttf_Subpanel5 = 152000;
mttr_Subpanel5 = 8;
mttf_JB1 = 522400;
mttr_JB1 = 8;
```

```
mttf_JB2 = 522400;
mttr_JB2 = 8;
mttf_JB3 = 522400;
mttr_JB3 = 8;
mttf_JB4 = 522400;
mttr_JB4 = 8;
mttf_JB5 = 522400;
mttr_JB5 = 8;
mttf_JB6 = 522400;
mttr_JB6 = 8;
mttf_PS1 = 215111;
mttr_PS1 = 8;
mttf_PS2 = 215111;
mttr_PS2 = 8;
mttf_PS3 = 215111;
mttr_PS3 = 8;
mttf_PS4 = 215111;
mttr_PS4 = 8;
mttf_PS5 = 215111;
mttr_PS5 = 8;
mttf_PS6 = 215111;
mttr_PS6 = 8;
mttf_PS7 = 215111;
mttr_PS7 = 8;
mttf_PS8 = 215111;
mttr_PS8 = 8;
mttf_PS9 = 215111;
mttr_PS9 = 8;
mttf_PS10 = 215111;
mttr_PS10 = 8;
RBD Model{
   block ACSource( MTTF = mttf_ACSource, MTTR = mttr_ACSource);
   block Generator( MTTF = mttf_Generator, MTTR = mttr_Generator);
   block Generator2( MTTF = mttf_Generator2, MTTR = mttr_Generator2);
   parallel p1(ACSource, Generator, Generator2);
   block ATS( MTTF = mttf_ATS, MTTR = mttr_ATS);
   block Subpanel1( MTTF = mttf_Subpanel1, MTTR = mttr_Subpanel1);
   block UPS1( MTTF = mttf_UPS1, MTTR = mttr_UPS1);
   block UPS2( MTTF = mttf_UPS2, MTTR = mttr_UPS2);
   parallel p4(UPS1, UPS2);
   block SDT1( MTTF = mttf_SDT1, MTTR = mttr_SDT1);
   block Subpanel2( MTTF = mttf_Subpanel2, MTTR = mttr_Subpanel2);
   series s6(SDT1, Subpanel2 );
   block SDT2(  MTTF = mttf_SDT2, MTTR = mttr_SDT2);
   block Subpanel3( MTTF = mttf_Subpanel3, MTTR = mttr_Subpanel3);
   series s7(SDT2, Subpanel3 );
   parallel p5(s6, s7);
```

```
    series s3(ATS, Subpanel1, p4, p5 );
    block ATS2( MTTF = mttf_ATS2, MTTR = mttr_ATS2);
    block Subpanel4( MTTF = mttf_Subpanel4, MTTR = mttr_Subpanel4);
    block UPS3( MTTF = mttf_UPS3, MTTR = mttr_UPS3);
    block SDT3( MTTF = mttf_SDT3, MTTR = mttr_SDT3);
    block Subpanel5( MTTF = mttf_Subpanel5, MTTR = mttr_Subpanel5);
    series s8(ATS2, Subpanel4, UPS3, SDT3, Subpanel5 );
    parallel p2(s3, s8);
    block JB1( MTTF = mttf_JB1, MTTR = mttr_JB1);
    block JB2( MTTF = mttf_JB2, MTTR = mttr_JB2);
    block JB3( MTTF = mttf_JB3, MTTR = mttr_JB3);
    parallel p9(JB1, JB2, JB3);
    block JB4( MTTF = mttf_JB4, MTTR = mttr_JB4);
    block JB5( MTTF = mttf_JB5, MTTR = mttr_JB5);
    block JB6( MTTF = mttf_JB6, MTTR = mttr_JB6);
    parallel p10(JB4, JB5, JB6);
    block PS1( MTTF = mttf_PS1, MTTR = mttr_PS1);
    block PS2( MTTF = mttf_PS2, MTTR = mttr_PS2);
    block PS3( MTTF = mttf_PS3, MTTR = mttr_PS3);
    parallel p11(PS1, PS2, PS3);
    block PS4( MTTF = mttf_PS4, MTTR = mttr_PS4);
    block PS5( MTTF = mttf_PS5, MTTR = mttr_PS5);
    block PS6( MTTF = mttf_PS6, MTTR = mttr_PS6);
    parallel p12(PS4, PS5, PS6);
    block PS7( MTTF = mttf_PS7, MTTR = mttr_PS7);
    block PS8( MTTF = mttf_PS8, MTTR = mttr_PS8);
    block PS10( MTTF = mttf_PS10, MTTR = mttr_PS10);
    series s14(PS8, PS10 );
    block PS9( MTTF = mttf_PS9, MTTR = mttr_PS9);
    parallel p13(PS7, s14, PS9);
    series s0(p1, p2, p9, p10, p11, p12, p13 );
    top s0;
    metric av = availability;
} main{ av = solve(Model, av);
      println(av);
      percentageDifference(
        model_ = "Model",
        metric_ = "av",
        samplingPoints = 5,
        parameters =  (
                        mttf_ACSource = [ 2190, 6570 ],
                        mttr_ACSource = [ 4, 12 ],
                        mttf_Generator = [ 1250, 3750 ],
                        mttr_Generator = [ 4, 12 ],
                        mttf_Generator2 = [ 1250, 3750 ],
                        mttr_Generator2 = [ 4, 12 ],
                        mttf_ATS = [ 12019, 36057 ],
```

```
                        mttr_ATS = [ 4, 12 ],
                        mttf_Subpanel1 = [ 76000, 228000 ],
                        mttr_Subpanel1 = [ 4, 12 ],
                        mttf_UPS1 = [ 25000, 75000 ],
                        mttr_UPS1 = [ 4, 12 ],
                        mttf_UPS2 = [ 25000, 75000 ],
                        mttr_UPS2 = [ 4, 12 ],
                        mttf_SDT1 = [ 70645, 211935 ],
                        mttr_SDT1 = [ 4, 12 ],
                        mttf_Subpanel2 = [ 76000, 228000 ],
                        mttr_Subpanel2 = [ 4, 12 ],
                        mttf_SDT2 = [ 70645, 211935 ],
                        mttr_SDT2 = [ 4, 12 ],
                        mttf_Subpanel3 = [ 76000, 228000 ],
                        mttr_Subpanel3 = [ 4, 12 ],
                        mttf_ATS2 = [ 12019, 36057 ],
                        mttr_ATS2 = [ 4, 12 ],
                        mttf_Subpanel4 = [ 76000, 228000 ],
                        mttr_Subpanel4 = [ 4, 12 ],
                        mttf_UPS3 = [ 25000, 75000 ],
                        mttr_UPS3 = [ 4, 12 ],
                        mttf_SDT3 = [ 70645, 211935 ],
                        mttr_SDT3 = [ 4, 12 ],
                        mttf_Subpanel5 = [ 76000, 228000 ],
                        mttr_Subpanel5 = [ 4, 12 ],
                        mttf_JB1 = [ 261200, 783600 ],
                        mttr_JB1 = [ 4, 12 ],
                        mttf_JB2 = [ 261200, 783600 ],
                        mttr_JB2 = [ 4, 12 ],
                        mttf_JB3 = [ 261200, 783600 ],
                        mttr_JB3 = [ 4, 12 ],
                        mttf_JB4 = [ 261200, 783600 ],
                        mttr_JB4 = [ 4, 12 ],
                        mttf_JB5 = [ 261200, 783600 ],
                        mttr_JB5 = [ 4, 12 ],
                        mttf_JB6 = [ 261200, 783600 ],
                        mttr_JB6 = [ 4, 12 ],
                        mttf_PS1 = [ 107555, 322666 ],
                        mttr_PS1 = [ 4, 12 ],
                        mttf_PS2 = [ 107555, 322666 ],
                        mttr_PS2 = [ 4, 12 ],
                        mttf_PS3 = [ 107555, 322666 ],
                        mttr_PS3 = [ 4, 12 ],
                        mttf_PS4 = [ 107555, 322666 ],
                        mttr_PS4 = [ 4, 12 ],
                        mttf_PS5 = [ 107555, 322666 ],
                        mttr_PS5 = [ 4, 12 ],
```

```
                         mttf_PS6 = [ 107555, 322666 ],
                         mttr_PS6 = [ 4, 12 ],
                         mttf_PS7 = [ 107555, 322666 ],
                         mttr_PS7 = [ 4, 12 ],
                         mttf_PS8 = [ 107555, 322666 ],
                         mttr_PS8 = [ 4, 12 ],
                         mttf_PS9 = [ 107555, 322666 ],
                         mttr_PS9 = [ 4, 12 ],
                         mttf_PS10 = [ 107555, 322666 ],
                         mttr_PS10 = [ 4, 12 ] ),
      output = ( type = "swing",
                        yLabel = "Steady-state availability",
                        baselineValue = av )); }
```

Listing C.4 – Mercury Script for RBD Sensitivity Analysis for Tier 4 Electrical Architecture.

```
mttf_ACSource1 = 4380;
mttr_ACSource1 = 8;
mttf_Generator1 = 2500;
mttr_Generator1 = 8;
mttf_Generator2 = 2500;
mttr_Generator2 = 8;
mttf_ACSource2 = 4380;
mttr_ACSource2 = 8;
mttf_ATS1 = 24038;
mttr_ATS1 = 8;
mttf_Subpanel1 = 152000;
mttr_Subpanel1 = 8;
mttf_UPS1 = 50000;
mttr_UPS1 = 8;
mttf_UPS2 = 50000;
mttr_UPS2 = 8;
mttf_SDT1 = 141290;
mttr_SDT1 = 8;
mttf_Subpanel2 = 152000;
mttr_Subpanel2 = 8;
mttf_SDT2 = 141290;
mttr_SDT2 = 8;
mttf_Subpanel3 = 152000;
mttr_Subpanel3 = 8;
mttf_ATS2 = 24038;
mttr_ATS2 = 8;
mttf_Subpanel4 = 152000;
mttr_Subpanel4 = 8;
mttf_UPS3 = 50000;
mttr_UPS3 = 8;
mttf_UPS4 = 50000;
mttr_UPS4 = 8;
```

```
mttf_SDT3 = 141290;
mttr_SDT3 = 8;
mttf_Subpanel5 = 152000;
mttr_Subpanel5 = 8;
mttf_JB1 = 522400;
mttr_JB1 = 8;
mttf_JB2 = 522400;
mttr_JB2 = 8;
mttf_JB3 = 522400;
mttr_JB3 = 8;
mttf_JB4 = 522400;
mttr_JB4 = 8;
mttf_JB5 = 522400;
mttr_JB5 = 8;
mttf_JB6 = 522400;
mttr_JB6 = 8;
mttf_JB7 = 522400;
mttr_JB7 = 8;
mttf_JB8 = 522400;
mttr_JB8 = 8;
mttf_PS1 = 215111;
mttr_PS1 = 8;
mttf_PS2 = 215111;
mttr_PS2 = 8;
mttf_PS3 = 215111;
mttr_PS3 = 8;
mttf_PS4 = 215111;
mttr_PS4 = 8;
mttf_PS5 = 215111;
mttr_PS5 = 8;
mttf_PS6 = 215111;
mttr_PS6 = 8;
mttf_PS7 = 215111;
mttr_PS7 = 8;
mttf_PS8 = 215111;
mttr_PS8 = 8;
mttf_PS9 = 215111;
mttr_PS9 = 8;
mttf_PS10 = 215111;
mttr_PS10 = 8;
mttf_PS11 = 215111;
mttr_PS11 = 8;
mttf_PS12 = 215111;
mttr_PS12 = 8;
RBD Model{
   block ACSource1( MTTF = mttf_ACSource1, MTTR = mttr_ACSource1);
   block Generator1( MTTF = mttf_Generator1, MTTR = mttr_Generator1);
```

```
block Generator2( MTTF = mttf_Generator2, MTTR = mttr_Generator2);
block ACSource2( MTTF = mttf_ACSource2, MTTR = mttr_ACSource2);
parallel p1(ACSource1, Generator1, Generator2, ACSource2);
block ATS1( MTTF = mttf_ATS1, MTTR = mttr_ATS1);
block Subpanel1( MTTF = mttf_Subpanel1, MTTR = mttr_Subpanel1);
block UPS1( MTTF = mttf_UPS1, MTTR = mttr_UPS1);
block UPS2( MTTF = mttf_UPS2, MTTR = mttr_UPS2);
parallel p4(UPS1, UPS2);
block SDT1( MTTF = mttf_SDT1, MTTR = mttr_SDT1);
block Subpanel2( MTTF = mttf_Subpanel2, MTTR = mttr_Subpanel2);
series s6(SDT1, Subpanel2 );
block SDT2(  MTTF = mttf_SDT2, MTTR = mttr_SDT2);
block Subpanel3( MTTF = mttf_Subpanel3, MTTR = mttr_Subpanel3);
series s7(SDT2, Subpanel3 );
parallel p5(s6, s7);
series s3(ATS1, Subpanel1, p4, p5 );
block ATS2( MTTF = mttf_ATS2, MTTR = mttr_ATS2);
block Subpanel4( MTTF = mttf_Subpanel4, MTTR = mttr_Subpanel4);
block UPS3( MTTF = mttf_UPS3, MTTR = mttr_UPS3);
block UPS4( MTTF = mttf_UPS4, MTTR = mttr_UPS4);
parallel p9(UPS3, UPS4);
block SDT3( MTTF = mttf_SDT3, MTTR = mttr_SDT3);
block Subpanel5( MTTF = mttf_Subpanel5, MTTR = mttr_Subpanel5);
series s8(ATS2, Subpanel4, p9, SDT3, Subpanel5 );
parallel p2(s3, s8);
block JB1( MTTF = mttf_JB1, MTTR = mttr_JB1);
block JB2( MTTF = mttf_JB2, MTTR = mttr_JB2);
block JB3( MTTF = mttf_JB3, MTTR = mttr_JB3);
parallel p10(JB1, JB2, JB3);
block JB4( MTTF = mttf_JB4, MTTR = mttr_JB4);
block JB5( MTTF = mttf_JB5, MTTR = mttr_JB5);
block JB6( MTTF = mttf_JB6, MTTR = mttr_JB6);
parallel p11(JB4, JB5, JB6);
block JB7( MTTF = mttf_JB7, MTTR = mttr_JB7);
block JB8( MTTF = mttf_JB8, MTTR = mttr_JB8);
parallel p12(JB7, JB8);
block PS1( MTTF = mttf_PS1, MTTR = mttr_PS1);
block PS2( MTTF = mttf_PS2, MTTR = mttr_PS2);
block PS3( MTTF = mttf_PS3, MTTR = mttr_PS3);
parallel p13(PS1, PS2, PS3);
block PS4( MTTF = mttf_PS4, MTTR = mttr_PS4);
block PS5( MTTF = mttf_PS5, MTTR = mttr_PS5);
block PS6( MTTF = mttf_PS6, MTTR = mttr_PS6);
parallel p14(PS4, PS5, PS6);
block PS7( MTTF = mttf_PS7, MTTR = mttr_PS7);
block PS8( MTTF = mttf_PS8, MTTR = mttr_PS8);
block PS9( MTTF = mttf_PS9, MTTR = mttr_PS9);
```

```
    parallel p15(PS7, PS8, PS9);
    block PS10( MTTF = mttf_PS10, MTTR = mttr_PS10);
    block PS11( MTTF = mttf_PS11, MTTR = mttr_PS11);
    block PS12( MTTF = mttf_PS12, MTTR = mttr_PS12);
    parallel p16(PS10, PS11, PS12);
    series s0(p1, p2, p10, p11, p12, p13, p14, p15, p16 );
    top s0;
    metric av = availability;
} main{ av = solve(Model, av);
       println(av);
       percentageDifference(
             model_ = "Model",
             metric_ = "av",
             samplingPoints = 5,
             parameters =  (
                         mttf_ACSource1 = [ 2190, 6570 ],
                         mttr_ACSource1 = [ 4, 12 ],
                         mttf_Generator1 = [ 1250, 3750 ],
                         mttr_Generator1 = [ 4, 12 ],
                         mttf_Generator2 = [ 1250, 3750 ],
                         mttr_Generator2 = [ 4, 12 ],
                         mttf_ACSource2 = [ 2190, 6570 ],
                         mttr_ACSource2 = [ 4, 12 ],
                         mttf_ATS1 = [ 12019, 36057 ],
                         mttr_ATS1 = [ 4, 12 ],
                         mttf_Subpanel1 = [ 76000, 228000 ],
                         mttr_Subpanel1 = [ 4, 12 ],
                         mttf_UPS1 = [ 25000, 75000 ],
                         mttr_UPS1 = [ 4, 12 ],
                         mttf_UPS2 = [ 25000, 75000 ],
                         mttr_UPS2 = [ 4, 12 ],
                         mttf_SDT1 = [ 70645, 211935 ],
                         mttr_SDT1 = [ 4, 12 ],
                         mttf_Subpanel2 = [ 76000, 228000 ],
                         mttr_Subpanel2 = [ 4, 12 ],
                         mttf_SDT2 = [ 70645, 211935 ],
                         mttr_SDT2 = [ 4, 12 ],
                         mttf_Subpanel3 = [ 76000, 228000 ],
                         mttr_Subpanel3 = [ 4, 12 ],
                         mttf_ATS2 = [ 12019, 36057 ],
                         mttr_ATS2 = [ 4, 12 ],
                         mttf_Subpanel4 = [ 76000, 228000 ],
                         mttr_Subpanel4 = [ 4, 12 ],
                         mttf_UPS3 = [ 25000, 75000 ],
                         mttr_UPS3 = [ 4, 12 ],
                         mttf_UPS4 = [ 25000, 75000 ],
                         mttr_UPS4 = [ 4, 12 ],
```

```
                        mttf_SDT3 = [ 70645, 211935 ],
                        mttr_SDT3 = [ 4, 12 ],
                        mttf_Subpanel5 = [ 76000, 228000 ],
                        mttr_Subpanel5 = [ 4, 12 ],
                        mttf_JB1 = [ 261200, 783600 ],
                        mttr_JB1 = [ 4, 12 ],
                        mttf_JB2 = [ 261200, 783600 ],
                        mttr_JB2 = [ 4, 12 ],
                        mttf_JB3 = [ 261200, 783600 ],
                        mttr_JB3 = [ 4, 12 ],
                        mttf_JB4 = [ 261200, 783600 ],
                        mttr_JB4 = [ 4, 12 ],
                        mttf_JB5 = [ 261200, 783600 ],
                        mttr_JB5 = [ 4, 12 ],
                        mttf_JB6 = [ 261200, 783600 ],
                        mttr_JB6 = [ 4, 12 ],
                        mttf_JB7 = [ 261200, 783600 ],
                        mttr_JB7 = [ 4, 12 ],
                        mttf_JB8 = [ 261200, 783600 ],
                        mttr_JB8 = [ 4, 12 ],
                        mttf_PS1 = [ 107555, 322666 ],
                        mttr_PS1 = [ 4, 12 ],
                        mttf_PS2 = [ 107555, 322666 ],
                        mttr_PS2 = [ 4, 12 ],
                        mttf_PS3 = [ 107555, 322666 ],
                        mttr_PS3 = [ 4, 12 ],
                        mttf_PS4 = [ 107555, 322666 ],
                        mttr_PS4 = [ 4, 12 ],
                        mttf_PS5 = [ 107555, 322666 ],
                        mttr_PS5 = [ 4, 12 ],
                        mttf_PS6 = [ 107555, 322666 ],
                        mttr_PS6 = [ 4, 12 ],
                        mttf_PS7 = [ 107555, 322666 ],
                        mttr_PS7 = [ 4, 12 ],
                        mttf_PS8 = [ 107555, 322666 ],
                        mttr_PS8 = [ 4, 12 ],
                        mttf_PS9 = [ 107555, 322666 ],
                        mttr_PS9 = [ 4, 12 ],
                        mttf_PS10 = [ 107555, 322666 ],
                        mttr_PS10 = [ 4, 12 ],
                        mttf_PS11 = [ 107555, 322666 ],
                        mttr_PS11 = [ 4, 12 ],
                        mttf_PS12 = [ 107555, 322666 ],
                        ttr_PS12 = [ 4, 12 ] ),
      output = ( type = "swing",
                       yLabel = "Steady-state availability",
                       baselineValue = av ));}
```

Listing C.5 – Mercury Script for RBD Sensitivity Analysis for Blade Server.

```
mttf_BaseBlade = 220000.0;
mttr_BaseBlade = 4;
mttf_CPU = 50000.0;
mttr_CPU = 4;
mttf_DIMMs = 480000.0;
mttr_DIMMs = 4;
mttf_HD = 200000.0;
mttr_HD = 4;
mttf_FiberChanelCard = 260000.0;
mttr_FiberChanelCard = 4;
mttf_EthernetCard = 1240000.0;
mttr_EthernetCard = 4;
RBD Model{
   block BaseBlade( MTTF = mttf_BaseBlade, MTTR = mttr_BaseBlade);
   block CPU( MTTF = mttf_CPU, MTTR = mttr_CPU);
   block DIMMs( MTTF = mttf_DIMMs, MTTR = mttr_DIMMs);
   block HD( MTTF = mttf_HD , MTTR = mttr_HD);
   block FiberChanelCard( MTTF = mttf_FiberChanelCard, MTTR =
    mttr_FiberChanelCard);
   block EthernetCard( MTTF = mttf_EthernetCard, MTTR = mttr_EthernetCard);
   series s0(BaseBlade, CPU, DIMMs, HD, FiberChanelCard, EthernetCard);
   top s0;
   metric av = availability;
} main{ av = solve(Model, av);
  println(av);
  percentageDifference(
            model_ = "Model",
            metric_ = "av",
            samplingPoints = 10,
            parameters =  (
                        mttf_BaseBlade = [110000, 330000],
                        mttr_BaseBlade = [ 4, 12 ],
                        mttf_CPU = [25000, 75000],
                        mttr_CPU = [ 4, 12 ],
                        mttf_DIMMs = [240000, 720000],
                        mttr_DIMMs = [ 4, 12 ],
                        mttf_HD = [100000, 300000],
                        mttr_HD = [ 4, 12 ],
                        mttf_FiberChanelCard = [130000, 390000],
                        mttr_FiberChanelCard = [ 4, 12 ],
                        mttf_EthernetCard = [620000, 1860000],
                        mttr_EthernetCard = [ 4, 12 ]),
      output = ( type = "swing",
                       yLabel = "Steady-state availability",
                       baselineValue = av ));}
```

Listing C.6 – Mercury Script for RBD Sensitivity Analysis for Chassis Blade.

```
mttf_Midplane = 310000.0;
mttr_Midplane = 4;
mttf_Software = 17520.0;
mttr_Software = 4;
mttf_PowerS = 670000.0;
mttr_PowerS = 4;
mttf_FiberSwitch = 320000.0;
mttr_FiberSwitch = 4;
mttf_EtherSwitch = 120000.0;
mttr_EtherSwitch = 4;
mttf_Blower = 620000.0;
mttr_Blower = 4;
mttf_B_Server = 27562.3350379806;
mttr_B_Server = 4;
RBD Model{
   block Midplane(MTTF = mttf_Midplane, MTTR = mttr_Midplane);
   block Software(MTTF = mttf_Software, MTTR = mttr_Software);
   koon PowerS(n = 4, k = 2, mttf = mttf_PowerS, mttr = mttr_PowerS);
   koon FiberSwitch(n = 2, k = 1, mttf = mttf_FiberSwitch, mttr =
   mttr_FiberSwitch);
   koon EtherSwitch(n = 2, k = 1, mttf = mttf_EtherSwitch, mttr =
   mttr_EtherSwitch);
   koon Blower(n=2, k = 1, mttf = mttf_Blower, mttr = mttr_Blower);
   koon B_Server(n=14, k = 6, mttf = mttf_B_Server, mttr = mttr_B_Server);
   series s0(Midplane, Software, PowerS, FiberSwitch, EtherSwitch, Blower,
   B_Server);
   top s0;
   metric av = availability;
} main{ av = solve(Model, av);
      println(av);
      percentageDifference(
            model_ = "Model",
            metric_ = "av",
            samplingPoints = 10,
            parameters =  (
                        mttf_Midplane = [155000, 465000],
                        mttr_Midplane = [ 4, 12 ],
                        mttf_Software = [8760, 26280],
                        mttr_Software = [ 4, 12 ],
                        mttf_PowerS = [335000, 1005000],
                        mttr_PowerS = [ 4, 12 ],
                        mttf_FiberSwitch = [160000, 480000],
                        mttr_FiberSwitch = [ 4, 12 ],
                        mttf_EtherSwitch = [60000, 180000],
                        mttr_EtherSwitch = [ 4, 12 ],
                        mttf_Blower = [310000, 930000],
```

```
                        mttr_Blower = [ 4, 12 ],
                        mttf_B_Server = [13781.16752, 41343.503],
                        mttr_B_Server = [ 4, 12 ]),
            output = ( type = "swing",
                       yLabel = "Steady-state availability",
                       baselineValue = av));}
```

Listing C.7 – Mercury Script for RBD Sensitivity Analysis for Tier 1 IT Architecture.

```
mttf_RT = 96154.0;
mttr_RT = 8;
mttf_QSFP = 1013999.9999999999;
mttr_QSFP = 4;
mttf_SW = 88684.6;
mttr_SW = 8;
mttf_UTP = 240000.0;
mttr_UTP = 4;
mttf_CR = 28849.0;
mttr_CR = 8;
RBD Model{
   block RT   ( MTTF = mttf_RT, MTTR = mttr_RT);
   block QSFP ( MTTF = mttf_QSFP, MTTR = mttr_QSFP);
   block SW   ( MTTF = mttf_SW, MTTR = mttr_SW);
   koon UTP( n = 15, k = 9, mttf = mttf_UTP, mttr = mttr_UTP);
   koon CR( n = 15, k = 9, mttf = mttf_CR, mttr = mttr_CR);
   series s0(RT, QSFP, SW, UTP, CR );
   top s0; metric av = availability;
} main{ av = solve(Model, av);
      println(av);
      percentageDifference(
            model_ = "Model",
            metric_ = "av",
            samplingPoints = 10,
            parameters =  (
      mttf_RT = [48077, 144231],
      mttr_RT = [ 4, 12 ],
      mttf_QSFP = [507000, 1521000],
      mttr_QSFP = [ 2, 6 ],
      mttf_SW = [44342, 133027],
      mttr_SW = [ 4, 12 ],
      mttf_UTP = [120000, 360000],
      mttr_UTP = [ 2, 6 ],
      mttf_CR = [14424, 43273],
      mttr_CR = [ 4, 12 ]),
            output = ( type = "swing",
                       yLabel = "Steady-state availability",
                       baselineValue = av));}
```

Listing C.8 – Mercury Script for RBD Sensitivity Analysis for Tier 2 IT Architecture.

```
mttf_RT = 96154.0;
mttr_RT = 8;
mttf_QSFP = 1013999.9999999999;
mttr_QSFP = 4;
mttf_QSFP2 = 1013999.9999999999;
mttr_QSFP2 = 4;
mttf_SW = 88684.6;
mttr_SW = 8;
mttf_SW2 = 88684.6;
mttr_SW2 = 8;
mttf_SFP = 980200.0000000001;
mttr_SFP = 4;
mttf_CR = 28849.099999999995;
mttr_CR = 8;
RBD Model{
   block RT    ( MTTF = mttf_RT, MTTR = mttr_RT);
   block QSFP  ( MTTF = mttf_QSFP, MTTR = mttr_QSFP);
   block QSFP2( MTTF = mttf_QSFP2, MTTR = mttr_QSFP2);
   parallel p1(QSFP, QSFP2);
   block SW    ( MTTF = mttf_SW, MTTR = mttr_SW);
   block SW2( MTTF = mttf_SW2, MTTR = mttr_SW2);
   parallel p2(SW, SW2);
   koon SFP( n = 30, k = 21, mttf = mttf_SFP, mttr = mttr_SFP);
   koon CR( n = 30, k = 21, mttf = mttf_CR, mttr = mttr_CR);
   series s0(RT, p1, p2, SFP, CR );
   top s0;
   metric av = availability;
} main{ av = solve(Model, av);
      println(av);
      percentageDifference(
            model_ = "Model",
            metric_ = "av",
            samplingPoints = 10,
            parameters =  (
                        mttf_RT = [48077, 144231],
                        mttr_RT = [ 4, 12 ],
                        mttf_QSFP = [507000, 1521000],
                        mttr_QSFP = [ 2, 6 ],
                        mttf_QSFP2 = [507000, 1521000],
                        mttr_QSFP2 = [ 2, 6 ],
                        mttf_SW = [44342, 133027],
                        mttr_SW = [ 4, 12 ],
                        mttf_SW2 = [44342, 133027],
                        mttr_SW2 = [ 4, 12 ],
                        mttf_SFP = [490100, 1470300],
                        mttr_SFP = [ 2, 6 ],
```

```
                        mttf_CR = [14424, 43273],
                        mttr_CR = [ 4, 12 ]),
            output = ( type = "swing",
                       yLabel = "Steady-state availability",
                       baselineValue = av));}
```

Listing C.9 – Mercury Script for RBD Sensitivity Analysis for Tier 3 IT Architecture.

```
mttf_RT = 96154.0;
mttr_RT = 8;
mttf_SFP = 980200.0000000001;
mttr_SFP = 4;
mttf_SW = 88684.6;
mttr_SW = 8;
mttf_UTP = 240000.0;
mttr_UTP = 4;
mttf_CR = 28849.099999999995;
mttr_CR = 8;
RBD Model{
   block RT( MTTF = mttf_RT, MTTR = mttr_RT);
   koon SFP( n = 3, k = 2, mttf = mttf_SFP, mttr = mttr_SFP);
   koon SW( n = 3, k = 2, mttf = mttf_SW, mttr = mttr_SW);
   koon UTP( n = 60, k = 48, mttf = mttf_UTP, mttr = mttr_UTP);
   koon CR( n = 60, k = 48, mttf = mttf_CR, mttr = mttr_CR);
   series s0(RT, SFP, SW, UTP, CR );
   top s0;
   metric av = availability;
} main{ av = solve(Model, av);
    println(av);
    percentageDifference(
              model_ = "Model",
              metric_ = "av",
              samplingPoints = 10,
              parameters =  (
                           mttf_RT = [48077, 144231],
                           mttr_RT = [ 4, 12 ],
                           mttf_SFP = [490100, 1470300],
                           mttr_SFP = [ 2, 6 ],
                           mttf_SW = [44342, 133027],
                           mttr_SW = [ 4, 12 ],
                           mttf_UTP = [120000, 360000],
                           mttr_UTP = [ 2, 6 ],
                           mttf_CR = [14424, 43273],
                           mttr_CR = [ 4, 12 ]),
            output = ( type = "swing",
                       yLabel = "Steady-state availability",
                       baselineValue = av));}
```

Listing C.10 – Mercury Script for RBD Sensitivity Analysis for Tier 4 IT Architecture.

```
mttf_RT = 96154.0;
mttr_RT = 8;
mttf_RT2 = 96154.0;
mttr_RT2 = 8;
mttf_SFP = 980200.0000000001;
mttr_SFP = 4;
mttf_SW = 88684.6;
mttr_SW = 8;
mttf_SFP_SW = 980200.0000000001;
mttr_SFP_SW = 4;
mttf_CR = 28849.099999999995;
mttr_CR = 8;
RBD Model{
   block RT (  MTTF = mttf_RT, MTTR = mttr_RT);
   block RT2(  MTTF = mttf_RT2, MTTR = mttr_RT2);
   parallel p1(RT, RT2);
   koon SFP( n = 4, k = 3, mttf = mttf_SFP, mttr = mttr_SFP);
   koon SW( n = 4, k = 3, mttf = mttf_SW, mttr = mttr_SW);
   koon SFP_SW( n = 120, k = 108, mttf = mttf_SFP_SW, mttr = mttr_SFP_SW);
   koon CR( n = 120, k = 108, mttf = mttf_CR, mttr = mttr_CR);
   series s0(p1, SFP, SW, SFP_SW, CR );
   top s0;
   metric av = availability;
}  main{ av = solve(Model, av);
    println(av);
    percentageDifference(
              model_ = "Model",
              metric_ = "av",
              samplingPoints = 10,
              parameters =  (
                            mttf_RT = [48077, 144231],
                            mttr_RT = [ 4, 12 ],
                            mttf_RT2 = [48077, 144231],
                            mttr_RT2 = [ 4, 12 ],
                            mttf_SFP = [490100, 1470300],
                            mttr_SFP = [ 2, 6 ],
                            mttf_SW = [44342, 133027],
                            mttr_SW = [ 4, 12 ],
                            mttf_SFP_SW = [490100, 1470300],
                            mttr_SFP_SW = [ 2, 6 ],
                            mttf_CR = [14424, 43273],
                            mttr_CR = [ 4, 12 ]),
                output = ( type = "swing",
                           yLabel = "Steady-state availability",
                           baselineValue = av));}
```