Optimum Utilization of Green Infrastructure in Cloud Data Centers

By

Eeman Fatima
(00000364862)

Supervisor: Assoc Prof Dr.Ayesha Maqbool

A thesis submitted to the faculty of Department of Computer Software Engineering, Military College of Signals, National University Of Sciences and Technology (NUST), Rawalpindi, in partial fulfillment of the requirements for a degree of MS in Computer Science.

(August, 2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by **Ms. Eeman Fatima,** Registration No **00000364862,** of **Military College of Signals** has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ____________________

Name of Supervisor **Assoc Prof Dr. Ayesha Maqbool**

Date: 04/9/24

Signature (HOD): Adnan A Khan

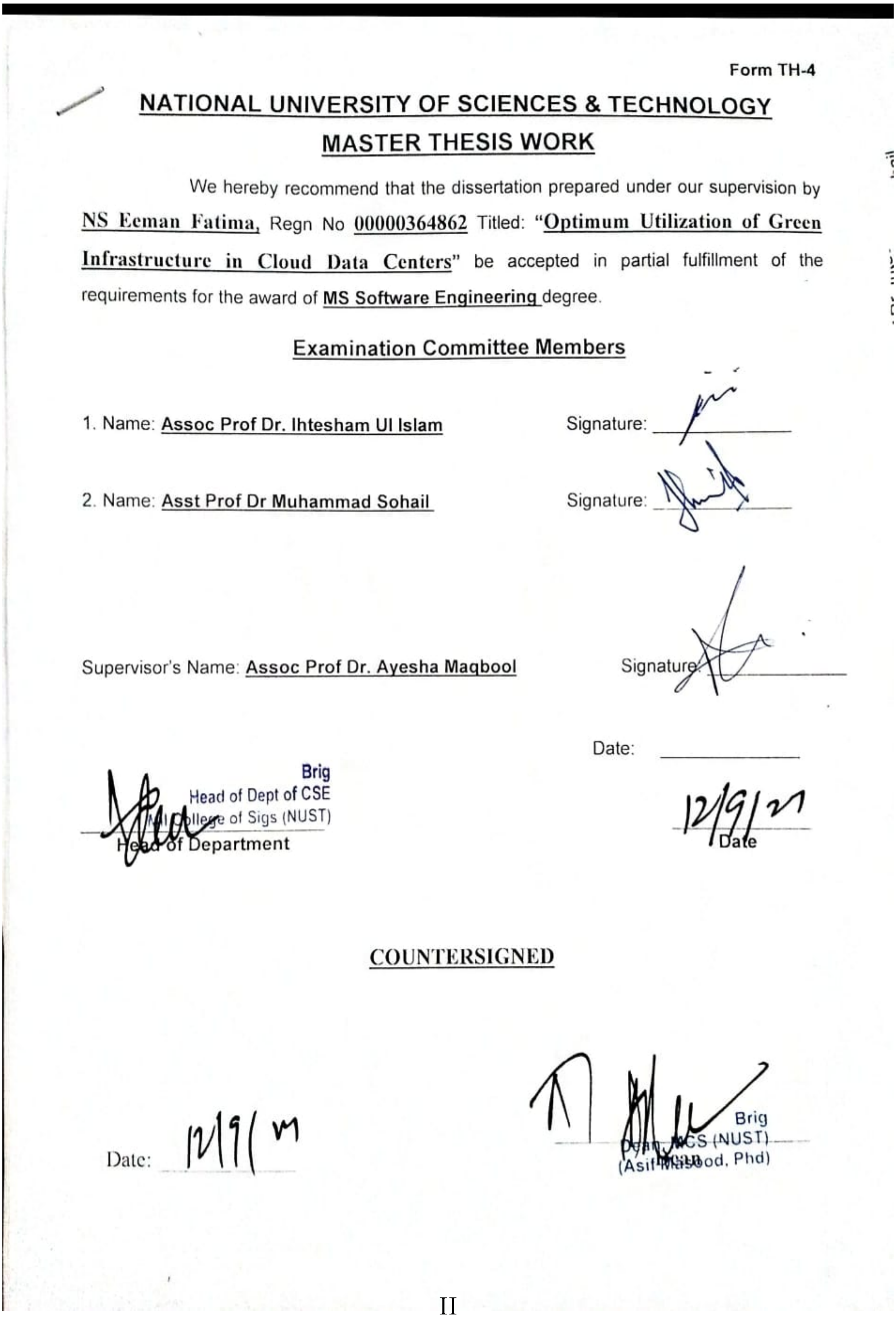
Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

Date: 12/9/24

Signature (Dean/Principal) Adnan A Khan

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

Date: 12/9/24



Form TH-4

NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by **NS Eeman Fatima,** Regn No **00000364862** Titled: "**Optimum Utilization of Green Infrastructure in Cloud Data Centers**" be accepted in partial fulfillment of the requirements for the award of **MS Software Engineering** degree.

Examination Committee Members

1. Name: **Assoc Prof Dr. Ihtesham Ul Islam** Signature: __________

2. Name: **Asst Prof Dr Muhammad Sohail** Signature: __________

Supervisor's Name: **Assoc Prof Dr. Ayesha Maqbool** Signature: __________

Date: __________

Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

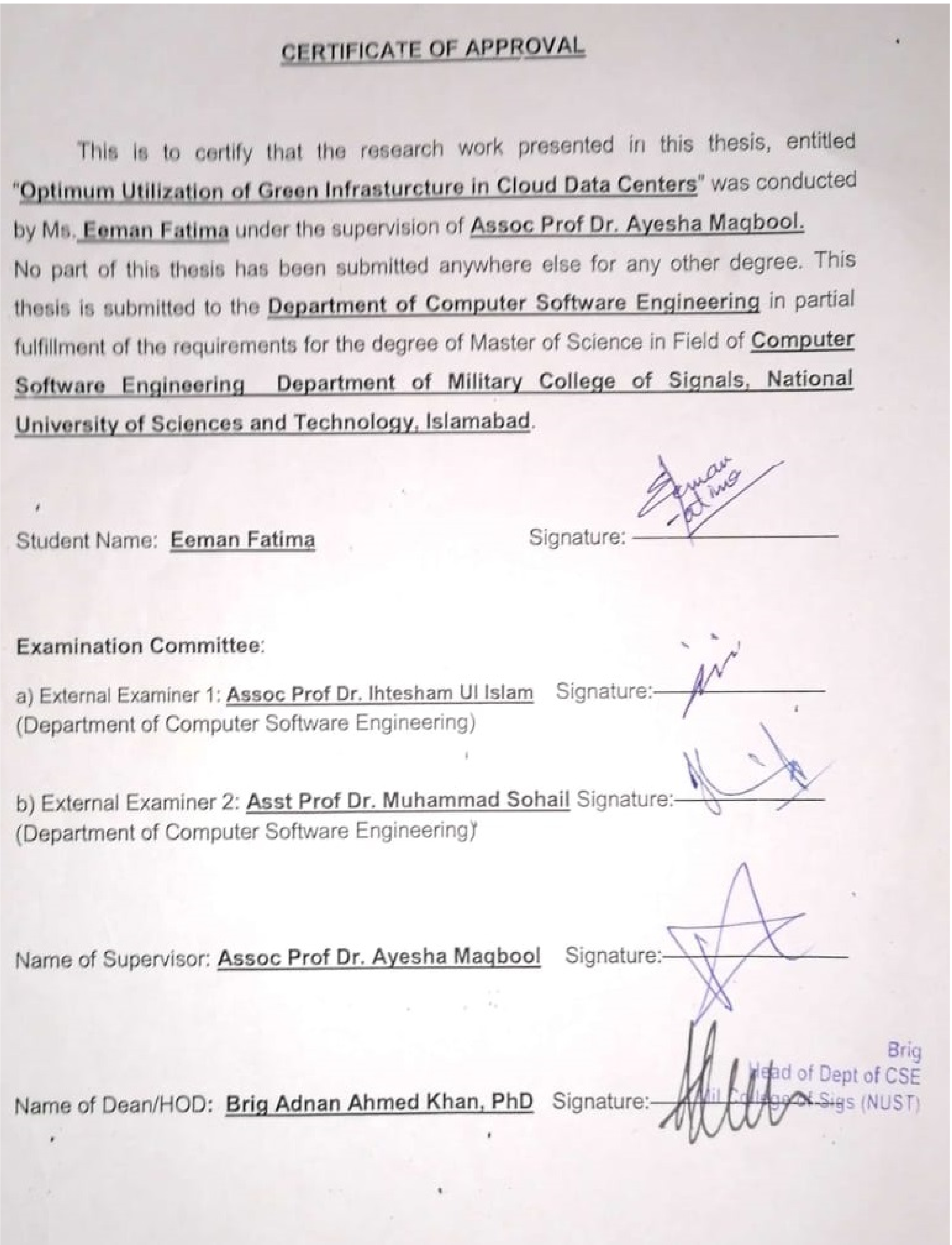
Head of Department

12/9/21
Date

COUNTERSIGNED

Date: 12/9/21

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)



CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "**Optimum Utilization of Green Infrasturcture in Cloud Data Centers**" was conducted by Ms. **Eeman Fatima** under the supervision of **Assoc Prof Dr. Ayesha Maqbool.**
No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Computer Software Engineering** in partial fulfillment of the requirements for the degree of Master of Science in Field of **Computer Software Engineering Department of Military College of Signals, National University of Sciences and Technology, Islamabad**.

Student Name: **Eeman Fatima** Signature: ____________

Examination Committee:

a) External Examiner 1: **Assoc Prof Dr. Ihtesham Ul Islam** Signature: ____________
(Department of Computer Software Engineering)

b) External Examiner 2: **Asst Prof Dr. Muhammad Sohail** Signature: ____________
(Department of Computer Software Engineering)

Name of Supervisor: **Assoc Prof Dr. Ayesha Maqbool** Signature: ____________

Name of Dean/HOD: **Brig Adnan Ahmed Khan, PhD** Signature: ____________

Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

AUTHOR'S DECLARATION

I, **Eeman Fatima**, hereby state that my MS thesis titled **"Optimum Utilization of Green Infrastructure in Cloud Data Centers"** is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Student Signature: Eeman Fatima

Name: **Eeman Fatima**

Date: 12-09-2024

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled **“Optimum Utilization of Green Infrastructure in Cloud Data Centers”** is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Eeman Fatima

Student Signature ______________

Name: **Eeman Fatima**

Date: 12-09-2024

DEDICATION

“In the name of Allah, the most Beneficent, the most Merciful"

I dedicate this thesis to my parents.

ACKNOWLEDGEMENTS

All praises to Allah for the strengths and His blessing in completing this thesis. I would like to convey my gratitude to my supervisor, Assoc. Prof. Dr. Ayesha Maqbool, PhD, for her supervision and constant support. Her priceless help of constructive comments and suggestions throughout the experimental and thesis works are major contributions to the success of this research. Also, I would thank committee members; Asst. Prof. Dr. Muhammad Sohail, and Assoc. Prof Dr. Ihtesham Ul Islam for their support and knowledge regarding this topic. Lastly, I am highly thankful to my parents for their constant support. I would like to thank them for their patience, cooperation and motivation in times of stress and hard work.

Abstract

With the recent surge in the use of cloud computing technology, the load on data centers has massively increased and so has the energy consumption to run these data centers. Data centers around the globe are huge consumers of energy and produce amounts of CO2 in return that is alarmingly dangerous for our environment. Globally several initiatives are being taken to curb the climate change and protect the environment from harmful gases that are released due to industrial activities and data center industry is also a part of it. There are several practices towards sustainable computing that are being promoted at wider scales and even the customer of this age prefer the industries that are aware of their contribution to a better world. Data center industry is already under huge influence of several incentives and standards that set the benchmarks for organizations to strive for green and sustainable solutions and this massive transformation is going to be more regulated in coming years as the demand for data centers and the awareness among the people is increasing rapidly. Against these changing trends the most viable argument is that the shift from traditional infrastructural practices to these new sustainable practices has huge capital, operational and recurrent costs. While these claims are true and pose a barrier to adoption of green solutions available in market yet it is also believed by the regulatory authorities that since the new trends are not only sustainable but also efficient so they benefit the organizations through massive savings on energy and operational costs thus produce a better ROI. The upfront costs are paid back to the companies through savings within few years and then they can hugely benefit from the savings and additional generated revenue. In this article we aim to find answers to the cost relevant barriers in way of green computing adoption for data centers.

Keywords: ROI, payback period, Green Data Centers, Green computing, sustainable, energy efficiency, economic efficiency

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

The need for sustainable practices and the increased awareness of environmental issues have made green data centers a prominent feature in the information technology landscape. Data centers produce CO2 because they use a lot of electricity, often generated from fossil fuels, to meet the rising demand for digital services. Also, the production and upkeep of their IT equipment add to their carbon emissions. Data centers and data transmission networks emitted around 330 Mt of CO2 equivalent in 2020, accounting for 0.9% of energy-related greenhouse gas emissions **[1]**. Data centers around the world use a significant amount of electricity, accounting for about 2% of the total global consumption as shown in Figure 1.1, which was around 240-340 terawatt-hours (TWh) in 2022 **[1]**. This consumption is expected to rise, potentially reaching 3-4% of global electricity by 2030. In the United States, data centers made up approximately 3% of the nation's total electricity use in 2022 **[2]**. Despite improvements in energy efficiency and the use of renewable energy, these emissions need to be halved by 2030 to align with global climate targets **[3]**. To align with the Sustainable Development Goals (SDGs), data centers should enhance energy efficiency and switch to more renewable energy sources, supporting clean energy (SDG 7:Affordable and Clean Energy) and also need to cut carbon emissions to fight climate change (SDG 13:Climate Action). By adopting green practices and technologies, data centers can promote responsible consumption and production (SDG 12: Responsible Consumption and Production) **[4]**.

Organizations can lay the groundwork for developing and running data centers with the least amount of negative environmental impact by comprehending approaches such as energy efficient hardware, renewable energy resources, cooling infrastructure optimization, virtualization and consolidation of servers, Life Cycle Assessment, waste heat reuse and building shell design can have a great impact on making data centers more sustainable. Google, Facebook, and Microsoft are leading the way in making data centers more eco-friendly by using renewable energy sources, advanced cooling techniques, and energy-efficient technologies. They are all committed to reducing their carbon footprints and improving sustainability.

Now the problem is that while the environmental benefits of green strategies in data centers often garner more attention, an equally important consideration is their economic sustainability. The overall performance and longevity of data center operations depend just as much on the economic effects of green initiatives as on their environmental impact and social responsibility of the corporate. The idea that using green strategies will cost a lot of money up front is one of the primary causes that organizations may prioritize the short-term financial effects over the long-term gains in the economy. There is not as much thorough research or bench marking studies that explicitly examine the financial advantages of green data center strategies. However, Researchers believe that sustainable data center solutions can lead to greater returns on investment by cutting energy costs, improving efficiency, and attracting eco-conscious clients. However, sophisticated financial modeling is needed to evaluate the economic impact of green strategies which must take into account variables like energy savings, gains in operational efficiency, and potential revenue from improved market positioning. Optimizing the utilization of infrastructure at a data center is crucial for efficiency, cost-effectiveness, and overall performance. Planning for future expansion and periodically evaluating the capacity requirements of infrastructure is another proactive strategy that guarantees efficient resource allocation and helps avoid bottlenecks such as sudden increase in demand. Also, employing monitoring tools to collect information on patterns in resource usage to spot possible capacity problems before they arise and using energy-saving hardware and cooling systems can cut down on power usage and running expenses.

Economic sustainability can be greatly enhanced by using sustainable solutions in data centers. In this sense, economic sustainability is the capacity to minimize adverse effects on the environment and society while preserving and enhancing economic performance over the long run. Organizations can enhance their overall economic performance and reduce their operational expenses by adopting green technologies, installing efficient cooling systems, and optimizing energy use. Effective data center management techniques, such virtualization and consolidation of servers, improve resource efficiency, lower the need for new infrastructure expenditures, and promote long-term financial viability. Investors and financial institutions are increasingly considering environmental, social, and governance (ESG) factors when making investment decisions. Businesses that adopt sustainable practices could have an easier time attracting capital and obtaining advantageous financing arrangements. This financial assistance provides the resources required for development and growth, which promotes economic sustainability.

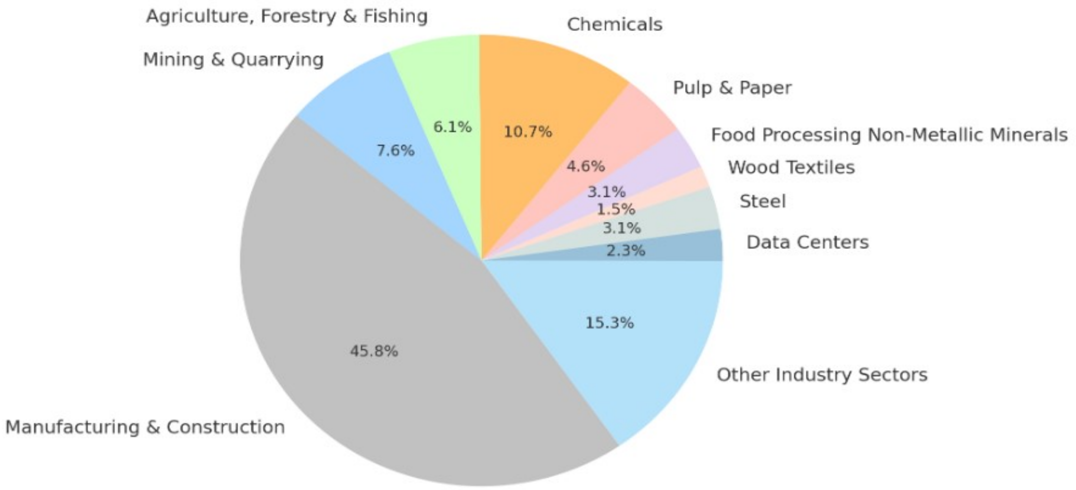


Figure 1.1: Power Consumption of Different Sectors

1.1.1 Data Center Physical Infrastructure

A data center is a type of computing facility where computer systems and related parts like storage and networking subsystems are kept. Typically, a data center offers data storage and computing resources to its clients. A data center's physical infrastructure encompasses all the hardware and facilities required to support its operations. This includes both the IT equipment and the non-IT equipment essential for maintaining optimal performance, security, and efficiency. Figure 1.2 shows all the software and hardware components of a data center including services, mechanical systems, IT system, electrical systems and connectivity and topology.

1.1.1.1 IT Equipment

Unsupervised Domain Adaptation (UDA) for Person Re-identification (Re-ID) is an important research area devoted to improving the performance of person re-identification models when applied to new, previously unknown domains. The basic goal of UDA is to transfer a model trained on a labelled source domain to a target domain devoid of labelled data. This modification is required to ensure that the model can recognize persons in a variety of real-world circumstances.

1. Servers: The primary computing units that process and store data.
2. Storage Systems: Devices like hard drives, SSDs, and storage area networks (SANs) that store data.

3. Networking Devices: Routers, switches, and firewalls that manage data traffic and ensure secure connectivity.
4. Cabling: Structured cabling systems that connect various devices within the data center.

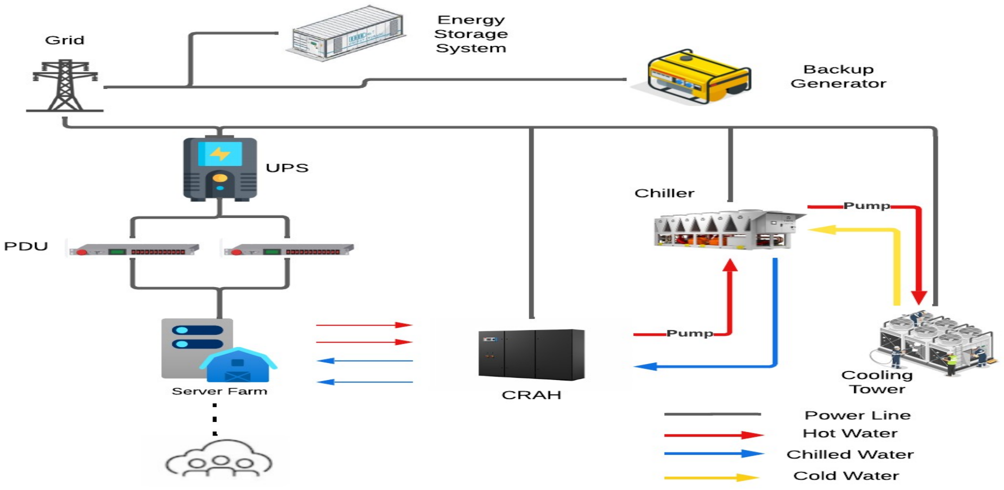


Figure 1.2: Data Center DCPI Connections

1.1.1.2 Non-IT Equipment

1. Power Distribution Units (PDUs): Distribute electrical power to servers and other IT equipment.
2. Uninterruptible Power Supplies (UPS): Provide backup power in case of electrical outages, ensuring continuous operation.
3. Generators: Offer extended backup power during prolonged power failures.
4. Cooling Systems: Include computer room air conditioning (CRAC) units, computer room air handling (CRAH) units, and chillers to maintain optimal temperature and humidity levels.
5. Racks: Structures that house servers, storage, and networking equipment in an organized manner.
6. Environmental Monitoring: Sensors and systems to monitor temperature, humidity, and other environmental factors to ensure optimal operating conditions.

7. . Security Systems: Physical security measures like access controls, surveillance cameras, and biometric scanners to protect against unauthorized access.

8. Fire Suppression Systems: Fire detection and suppression systems to protect the data center from fire hazards.

9. Building Infrastructure: Includes raised floors for efficient air circulation, perforated tiles, false ceilings for air distribution, and robust physical structures to support the entire data center.

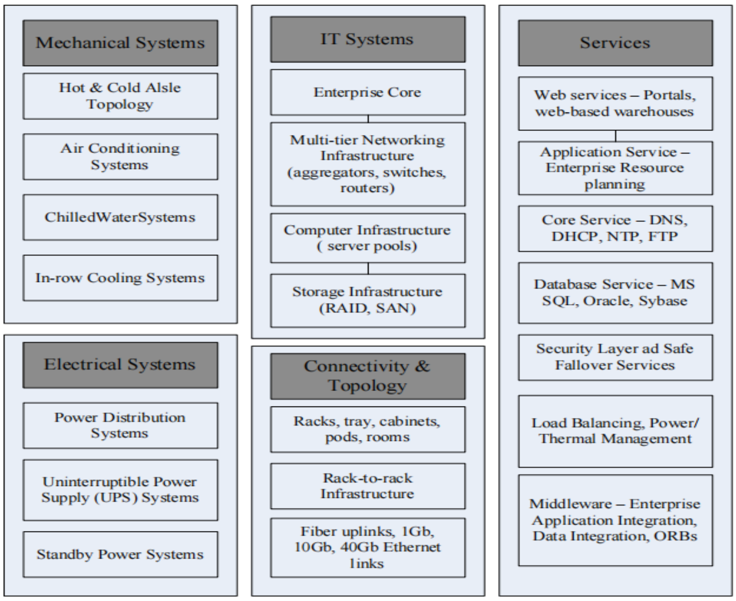


Figure 1.3: Components of a Data Center

1.1.2 Green Data Center Strategies and Practices

In my previous paper "Data Centers Sustainability: Approaches to Green Data Centers" **[1]** published in ComTech, 2023 we discussed several green strategies to enhance the sustainability of data centers. These strategies are crucial for reducing the environmental impact of data centers, which are significant consumers of energy and contributors to carbon emissions. The main strategies discussed are renewable energy, energy-efficient servers, efficient cooling mechanisms, and green building infrastructure. These strategies are discussed further below:

1.1.2.1 Renewable Energy

The use of renewable energy resources is emphasized as a pivotal strategy for reducing the carbon footprint of data centers **[2]**. Data centers can be powered by green energy sources such as solar, wind, hydro, and geothermal power. The paper **[3]** highlights the importance of both on-site and off-site renewable energy generation, with examples like Google's initiative to use geothermal energy for its Nevada data center and the replacement of diesel generators with large batteries in Belgium. The deployment of renewable energy solutions along with adequate battery storage can help achieve carbon neutrality. Different renewable energy sources are given in Table 1.1.

Table 1.1. Widely Used Renewable Energy Resources

Serial Number	**Renewable Energy Source**	**Description**	**Key Benefits**
1	Solar Energy	Energy harnessed from sunlight using photovoltaic cells or solar panels.	Clean, abundant, reduces electricity bills, low maintenance costs.
2	Wind Energy	Energy generated from wind using turbines.	Low operational costs, renewable and sustainable, reduces greenhouse gas emissions.
3	Hydro Energy	Energy produced from moving water, typically in dams.	Reliable, low emissions, provides water storage for agriculture.
4	Geothermal Energy	Energy derived from the heat within the Earth's crust.	Stable and reliable, low emissions, small land footprint.
5	Biomass Energy	Energy obtained from organic materials like plant and animal waste.	Reduces waste, carbon neutral, diverse fuel supply.

1.1.2.2 Energy Efficient Servers

Energy efficiency in servers is crucial as servers are the primary consumers of power in data centers. **[4]** discusses strategies such as virtualization, which allows multiple virtual machines to run on a single physical server, thus optimizing resource use and reducing energy consumption.

It also highlights the development of green resource management techniques and adaptive fault tolerance to improve server efficiency. Energy-efficient servers incorporate several advanced technologies to reduce power consumption and improve performance. Virtualization allows multiple virtual machines to run on a single physical server, optimizing resource use and significantly cutting down energy usage. Modern servers also employ dynamic power management, which adjusts power consumption based on workload demands, ensuring minimal energy wastage during low-activity periods. Innovations like solid-state drives (SSDs) over traditional hard drives have improved energy efficiency due to lower power requirements and faster data access speeds. Additionally, newer server designs focus on better thermal management with improved airflow and liquid cooling systems, reducing the need for excessive air conditioning and thereby saving energy.

1.1.2.3 Efficient Cooling Mechanisms

Cooling mechanisms are another major focus area as they significantly contribute to the overall energy consumption of data centers.

1. Natural Cooling: Leveraging natural cold environments, such as Microsoft's underwater data center project and Facebook's data center in Sweden.
2. Liquid Cooling: Using hot and cold water systems, which have higher heat transfer capacity than air cooling.
3. Rack Arrangement: Implementing configurations like hot-aisle/cold-aisle to optimize cooling efficiency.
4. Free Cooling: Utilizes the external environment's cool air to aid in cooling data centers, reducing the need for mechanical refrigeration. Techniques include air-side and water-side economization.
5. Adiabatic Cooling: Lowers the temperature of air through the evaporation of water. It is particularly effective in dry climates and can significantly reduce the energy needed for traditional air conditioning systems.
6. Immersion Cooling: Involves submerging servers in a thermally conductive, but electrically insulatingliquid. This method allows for direct cooling of components, leading to highly efficient heat removal.
7. Phase Change Cooling: Uses phase change materials (PCMs) that absorb heat as they change from solid to liquid, providing efficient thermal management and reducing reliance on traditional cooling systems.

8. Chilled Beam Cooling: Employs chilled water running through ceiling beams to absorb heat from the room. This method provides efficient cooling with reduced energy consumption compared to traditional air conditioning systems.
9. Thermal Energy Storage: Involves storing energy in the form of ice or chilled water during off-peak hours and using it for cooling during peak load times, thereby reducing the load on cooling systems and enhancing energy efficiency.

1.1.2.4 Green Building Infrastructure

Green building infrastructure includes the design and construction of data centers with sustainable materials and practices. Our previous paper discusses the importance of using eco-friendly construction materials, installing thermal insulators, and designing facilities with raised floors and false ceilings to enhance thermal efficiency. Certifications like LEED (Leadership in Energy and Environmental Design) are mentioned as benchmarks for achieving and recognizing sustainable building practices. Green building infrastructure practices for data centers focus on optimizing energy use and enhancing sustainability. Utilizing perforated tiles and raised floors is essential for efficient air circulation, allowing cool air to flow directly to the servers and hot air to be expelled, thus reducing the load on cooling systems. False ceilings further aid in managing airflow and maintaining optimal temperatures by creating a plenum space for air distribution. Incorporating energy-efficient materials such as recycled steel and sustainable insulation can significantly reduce the building's carbon footprint. Additionally, using low-VOC paints and sustainable construction materials ensures minimal environmental impact. Implementing advanced HVAC systems, like high-efficiency heating and cooling units, coupled with renewable energy sources such as solar panels, contributes to a sustainable and energy-efficient data center infrastructure. These practices not only reduce energy consumption but also enhance the longevity and efficiency of the data center, making it more environmentally friendly.

1.1.3 Problem Statement

The rapid growth of data centers has raised environmental concerns due to their high carbon emissions and energy consumption. Despite the potential of green practices to address these issues, their adoption is hindered by perceived economic barriers and a lack of clear ROI analysis. This research aims to conduct a comprehensive economic analysis of green practices in data centers, evaluating the financial viability and ROI of sustainable technologies such as renewable energy, efficient cooling systems, and waste heat recovery. By providing robust data and insights, the study seeks to demonstrate that green strategies offer both environmental benefits and attractive investment opportunities, ultimately encouraging increased investment in sustainable data center technologies to significantly reduce their carbon footprint.

1.1.4 Reason/Justification

The demand for data centers is increasing rapidly with the rapid development of the ICT industry and more and more users getting access to the technology. The negative environmental impacts of data centers, starting from their deployment to their running/operational phase, cannot be ignored. Data centers contribute to CO2 emissions globally harming the environment in a very negative way and to curb these emissions to preserve the environment, various strategies to make computing green are implemented. But there is a common misconception that these strategies cost a lot if adopted. The justification against this misconception is that if estimated annually, sustainable solutions can increase the revenue generated for the cloud providers by generating more return on investment compared to non-green data centers for cloud. This way the operational and deployment costs are compensated within a few years and more profit can be achieved. This helps achieve both environmental and economic sustainability for cloud organizations.

1.1.5 Objectives

Convincing organization management teams to approve hardware refreshes is challenging without a solid business plan that clearly outlines when a return on investment (ROI) will be realized. This study focuses on providing different cost and energy models for cloud deployment if the identified strategies for green computing are implemented which would be done in following ways:

- Identify all the available strategies through which green computing in the cloud can be achieved and potential areas where the green technology can be deployed.
- Estimate cost of deployment of the identified strategies and operational costs that the maintenance of the technology may have.
- Compare these costs with the costs of traditional cloud infrastructure that does not have any implementation of green strategies.
- Formulate a cost model to compare green infrastructure vs the traditional infrastructure of cloud.
- Find Return on Investment; plan and devise recommendations based on that plan for decision makers to choose green options over others.

As shown in Figure 4 our main objective is to find the payback period point after initial investment in green data centers has been made after which the savings through efficient equipment accumulates into the ROI and generates revenue for data centers.

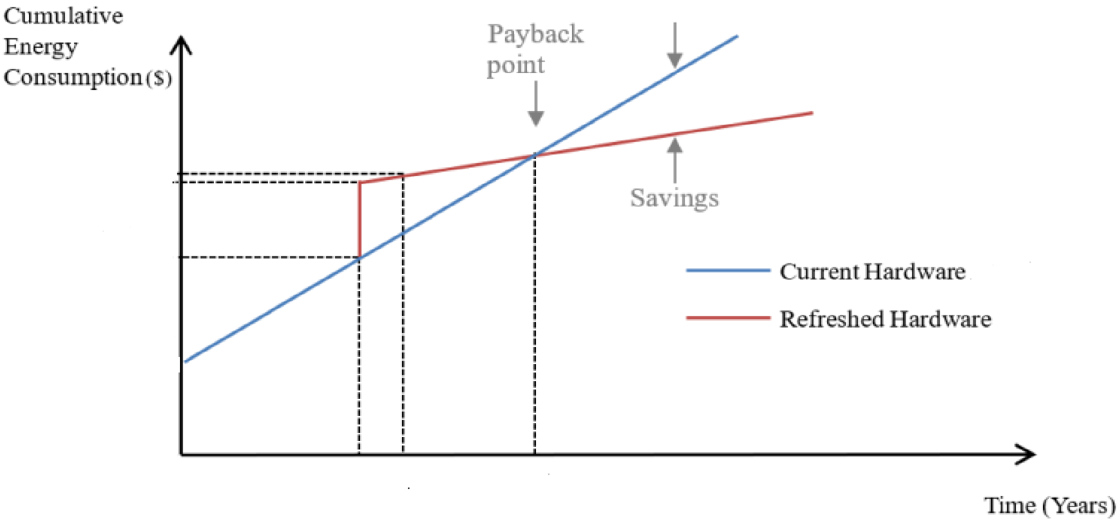


Figure 1.4: Cumulative Energy Consumption of Existing system vs Energy Efficient System

1.1.6 Relevance to National Needs

Pakistan is going through energy, financial and climate crisis. With the privacy concerns getting enough attention today, it can be predicted that in near future data localization policies could be proposed and implemented within Pakistan as well and soon big social media sites and other companies for example Google and Microsoft could be asked to locate data storage and other cloud services within Pakistan. For that purpose, deployment of huge data centers will be the first step. For the ongoing energy, climate and financial crisis all over the world and within Pakistan, application of sustainable solutions is of foremost priority that can help in terms of reducing the cost, curbing negative environmental impacts of technology and utilization of least energy possible. It will help decision makers with help of recommendations to make better decisions when it comes to green data centers.

1.1.7 Area of Application

This research can be applied in cloud data centers to promote the sustainable solutions available today and in future in a more optimized manner and measuring investments in such businesses for producing a good ROI for data center owners.

1.1.8 Advantages

This research will have following advantages in the ICT sector:

1. Identification of different sustainable solutions for Cloud Data Centers.
2. Utilization of sustainable technologies in Data centers modeled in an optimized way.
3. Recommendations for making data centers more sustainable both economically and environmentally.
4. Recommendation and guidance for investors to motivate them to invest in green data center technological solutions
5. Provide a proof that ROI generated through deployment of green infrastructure is more than that of traditional infrastructure and it is a more suitable option for a country like Pakistan which is going through an energy and financial crisis.

1.2 Thesis Outline

This thesis unfolds through a structured series of chapters, each dedicated to specific aspects of green data centers and their associated costs.

Chapter 1: This is the introductory chapter that highlights reasons of conducting this research, outlines objectives, and areas of application

Chapter 2: This section provides extensive literature review in the field of green data centers, energy efficient servers, cooling infrastructure and renewable energy sources

Chapter 3: In this chapter, various power models for different data center equipment are discussed and financial analysis metrics for data centers are discussed.

Chapter 4: Gives insights of various certifications, frameworks and benchmark programs for sustainable data centers.

Chapter 5: In this chapter, we discuss how different power models are applied to calculate total power consumed by different components of data center. We also see how different data center hardware is refreshed to more efficient hardware and how it effects the total energy consumption and ROI.

Chapter 6: This concluding chapter gives final remarks and reflections on this research.

Chapter 2

Literature Review

2.1 Background and Significance

The study **[5]** explores the role of green computing in reducing the costs of modern industrial products. The study emphasizes the environmental and economic impacts of increasing technical waste, which includes toxic materials and heavy plastics that pose significant environmental hazards. The findings demonstrate that implementing these practices can reduce the cost of direct raw materials by 42% and lower the total cost of manufacturing a smartphone by 40%. The study also discusses the historical context of green computing, tracing its origins to the Energy Star program in 1992, which aimed to increase energy efficiency in electronic devices. Additionally, the research underscores the economic benefits of green computing, including increased profitability, competitive pricing, and improved resource utilization. The methodology involves a detailed cost analysis using activity-based costing (ABC) to quantify the savings from recycling and reuse strategies. Key performance indicators (KPIs) for green data centers include the Service Level Agreement (SLA) and, crucially, the overall reduction in energy consumption. The benefits of a green data center go beyond just cutting energy costs; they also extend the lifespan of equipment, lower maintenance expenses, protect the environment, and reduce the strain on the electricity grid, among other advantages.

2.1.1 Energy Consumption of Cloud Data Center

Authors in **[6]** present an optimization framework for integrating renewable energy into distributed data centers while ensuring high availability and minimizing total cost of ownership (TCO). The research addresses the dual challenges of maintaining fault tolerance through spare capacity provisioning and optimizing the operating costs by incorporating both green and brown energy sources. The results indicate that incorporating up to 40% green energy is feasible with only a marginal increase in TCO compared to traditional cost-aware models. The study underscores the importance of considering both energy costs and environmental impacts in data center operations. Additionally, it highlights the significance of strategic location selection to avoid simultaneous failures and leverage uncorrelated renewable energy sources. Authors in **[7]** investigate the energy consumption dynamics in cloud computing, comparing it with conventional computing to identify potential energy savings. It evaluates the energy efficiency of these models by analyzing power consumption in data processing, data storage, and data transport. The findings reveal that cloud computing can offer significant energy savings through techniques like virtualization and server consolidation, particularly for low-intensity and infrequent computing tasks. However, the study also notes that cloud computing may consume more energy than conventional local computing under certain conditions. The research highlights the trade-offs between energy savings and performance, suggesting that cloud computing is not always the most environmentally friendly option. **[8]**, **[9]** offer comprehensive insights into the energy consumption in cloud computing data centers and strategies for energy and cost savings. The primary factors affecting energy consumption include servers, cooling systems, power conditioning systems, and networking equipment. Servers consume a significant portion of energy, even in idle states, due to their need to remain operational and ready for workloads. Cooling systems, essential for maintaining optimal operating temperatures, also contribute significantly to energy usage, with the Computer Room Air Handler (CRAH) units being particularly energy-intensive. Various models and techniques have been proposed to optimize energy use, such as dynamic voltage/frequency scaling (DVFS) and virtualization, which allow for better utilization of resources by consolidating multiple virtual machines on fewer physical servers. Power conditioning systems, including Uninterruptible Power Supplies (UPS) and Power Distribution Units (PDU), add to the energy consumption due to their inherent inefficiencies and energy loss during power conversions. To improve energy efficiency, strategies like intelligent cooling management, improved server utilization, and the adoption of advanced cooling technologies are critical. These strategies not only reduce energy costs but also minimize the environmental impact of data centers, contributing to sustainable cloud computing practices.

2.1.2 Energy Efficient Servers

About 20% of data center servers, according to the Uptime Institute, are out-of-date, and almost 50% of data centers don't have regular procedures in place to find and discard these underutilized servers. Moreover, more than 50% of senior executives working in these data centers think that less than 5% of their servers are "comatose," or not in use. The authors in **[10]** addresses the growing energy consumption in cloud data centers, presenting a distributed energy consumption measurement system (DEM) for heterogeneous cloud environments. The study proposes a multi-component power

Table 2.1. Power Distribution in a Data Center

Input	**Equipment**	**Distribution (%)**
Electrical Power In: 100%	Chiller	23%
	Humidifier	3%
	CRAC/CRAH (Cooling)	15%
	IT Equipment	47%
	PDU	3%
	UPS	6%
	Lighting	2%
	Generator	1%

model to estimate the energy consumption of key components such as CPU, memory, and disk. The DEM system can measure energy consumption with high accuracy in both Linux and Windows environments. The paper highlights the inefficiencies in data center energy usage and the need for improved energy management to reduce operational costs and environmental impact. Experimental results confirm the system's accuracy in estimating cluster energy consumption, providing a valuable tool for energy optimization in cloud data centers. **[11]** focuses on developing a cloud server energy consumption measurement system tailored for heterogeneous cloud environments. It introduces a power model for servers and investigates the mathematical relationship between resource usage of key components and system energy consumption. The study highlights the importance of accurate energy measurement for efficient energy management in cloud data centers. The proposed system supports various CPU power models and introduces an I/O-mode aware disk power model for precise energy calculation. The findings demonstrate that the system can accurately measure and estimate energy consumption, offering a practical solution for monitoring and optimizing energy usage in diverse cloud environments. In **[12]** authors investigates the power and energy consumption of database servers, focusing on the significant impact of electrical energy costs on the performance of database management systems (DBMS). The study introduces an empirical methodology using multiple-linear regression and factorial experimental design to create cost models for predicting power and energy demands based on readily available statistics like selectivity factors and tuple sizes. This methodology does not require individual hardware component measurements, providing an accurate estimation of power and energy consumption. The results show that the proposed models are more precise than alternative methods, suggesting that they can be effectively used for provisioning decisions and energy-aware query optimization in data centers.

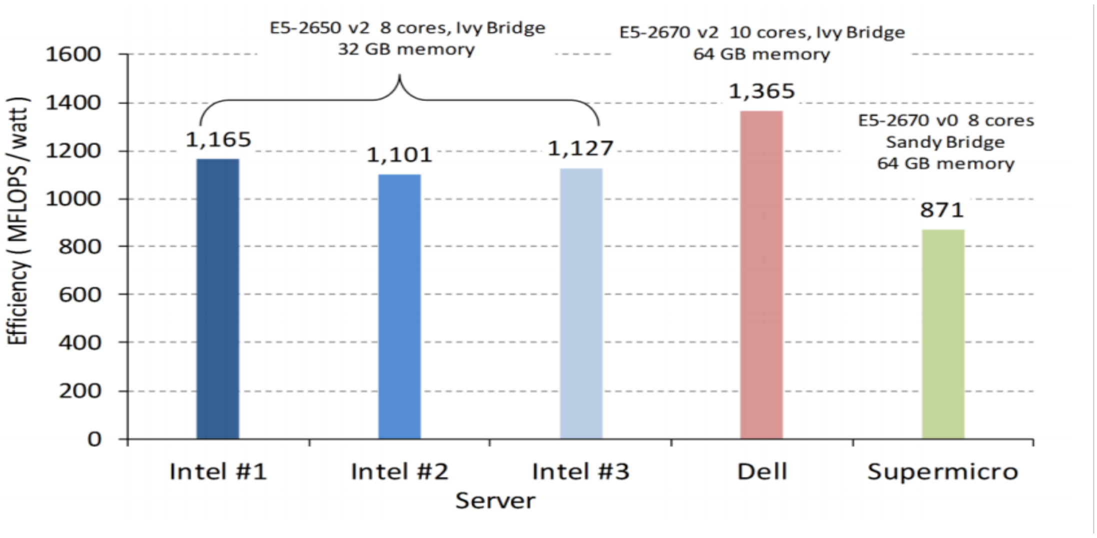


Figure 2.1: Compute Efficiency of Different Intel and Dell Servers

2.1.3 Efficient Cooling

Data centers, as the backbone of the digital era, consume substantial amounts of energy, primarily due to the need for efficient cooling systems to maintain optimal operating conditions for IT equipment. With the rising concerns over energy consumption and environmental impact, innovative cooling technologies have emerged as pivotal solutions. This literature review explores various energy-efficient cooling technologies, such as free cooling and air cooling, and discusses their cost-effectiveness**[13]**. In the past, data centers relied on chillers for cooling, but these systems consumed a lot of energy, contributing to high Power Usage Effectiveness (PUE). To reduce their PUE, green data centers are now adopting alternative cooling technologies and continually researching advanced options **[43]**.

2.1.3.1 Free Cooling

Free cooling, also known as economizer cooling, leverages external environmental conditions to reduce the need for traditional mechanical cooling. This method capitalizes on naturally cool air or water from the external environment to dissipate heat within the data center **[14]**.

2.1.3.2 Air-Side Economizers

Air-side economizers draw cool outside air into the data center and expel warm air, significantly reducing the reliance on energy-intensive chillers. This approach is particularly effective in cooler climates, where ambient air temperatures frequently fall below the required indoor temperature and is highly energy saving **[15]**.

2.1.3.3 Air Cooling

Air cooling remains one of the most widely used methods due to its simplicity and cost-effectiveness. Innovations in air cooling have led to significant improvements in energy efficiency.

2.1.3.4 Direct Air Cooling

Direct air cooling involves circulating ambient air directly through the IT equipment. The major advantage of this method is its low initial cost and ease of implementation. However, its efficiency heavily depends on the cleanliness and temperature of the outside air.

The cost-effectiveness of these cooling technologies extends beyond mere energy savings. Initial investment, operational costs, maintenance, and potential scalability play crucial roles in evaluating their financial viability.Free cooling systems, particularly those involving water-side economizers, often require substantial initial investments due to the need for additional infrastructure, such as cooling towers or heat exchangers. In contrast, air cooling systems generally have lower upfront costs, making them accessible for a wider range of data centers, especially smaller facilities but are less efficient **[16]**. Table 2.2 lists various cooling mechanisms used in data centers.

2.1.4 Efficiency Metrics

Efficiency metrics for data centers are crucial for assessing their performance, resource utilization, and overall operational efficiency. These metrics help data center operators optimize their infrastructure, reduce costs, and improve sustainability. Here are some key efficiency metrics for data centers in Table 2.3.

Table 2.2. Cooling Mechanisms in Data Centers

Cooling System	**Description**	**Effectiveness**	**Efficiency**	**Cost**
CRAC (Computer Room Air Conditioning)	Direct expansion refrigeration cooling.	Good for small to medium data centers.	Moderate; energy-intensive.	Moderate to High.
CRAH (Computer Room Air Handling)	Uses chilled water for cooling.	Highly effective for large data centers.	Higher efficiency than CRAC.	High initial, better long-term savings.
Liquid Cooling	Coolants absorb heat directly from components.	Extremely effective for high-density setups.	Very high efficiency.	High initial, lower operational costs.
In-Row and In-Rack Cooling	Positioned close to servers.	Very effective for targeted cooling.	Highly efficient.	Moderate to High.
Heat Rejection Systems	Expels heat externally via cooling towers or coolers.	Effective in temperature control.	Varies; can be optimized.	Moderate to High.
Airflow Management	Manages airflow direction and quality.	Enhances other cooling methods.	Improves overall efficiency.	Low to Moderate.
Free Cooling	Utilizes outside air during cool weather.	Highly effective in suitable climates.	Very efficient.	Low to Moderate.
Adiabatic Cooling	Uses water evaporation to cool air.	Effective in dry climates.	Efficient; reduces mechanical cooling need.	Moderate initial, lower operational costs.

Table 2.3. DATA CENTERS METRICS

Metric	**Description**	**Formula**
PUE (Power Usage Effectiveness)	Measures the overall energy efficiency of a data center by comparing total facility energy use to IT equipment energy use.	$\text{PUE} = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}}$
pPUE (Partial Power Usage Effectiveness)	Evaluates energy efficiency in specific zones within a data center.	$\text{pPUE} = \frac{\text{Energy for Specific Zone}}{\text{IT Equipment Energy in Zone}}$
DCiE (Data Center Infrastructure Efficiency)	Inverse of PUE, expressed as a percentage to show how much of the total energy is used by IT equipment.	$\text{DCiE} = \frac{1}{\text{PUE}} \times 100\%$
CUE (Carbon Usage Effectiveness)	Measures the amount of CO_2 emissions produced per unit of IT equipment energy used.	$\text{CUE} = \frac{\text{Total CO}_2\text{ Emissions}}{\text{IT Equipment Energy}}$
WUE (Water Usage Effectiveness)	Assesses water conservation efficiency by measuring water use per unit of IT equipment energy.	$\text{WUE} = \frac{\text{Total Water Usage}}{\text{IT Equipment Energy}}$
EDE (Electronics Disposal Efficiency)	Evaluates how effectively a data center recycles and disposes of electronic waste.	$\text{EDE} = \frac{\text{Recycled Electronics}}{\text{Total Electronics Disposed}}$
CoP (Coefficient of Performance)	Measures the efficiency of cooling systems by comparing cooling output to the energy input required.	$\text{CoP} = \frac{\text{Cooling Output}}{\text{Energy Input}}$
Performance per Watt	Compares server performance to power consumption to evaluate energy efficiency of computing performance.	$\text{Performance per Watt} = \frac{\text{Computing Performance}}{\text{Power Consumption}}$
SPEC Power Benchmark	Standardized test for measuring server energy efficiency under different workloads.	SPECpower_ssj2008

PUE is the most widely used metric for data centers. The nearer it is to 1 the more efficient a data center is. Different PUE measurements and their efficiency is given in table 2.4.

Table 2.4. PUE and DCiE Efficiency Measurements

PUE	**Efficiency**	**DCiE**
3.0	Very Inefficient	33%
2.5	Inefficient	40%
2.0	Average	50%
1.5	Efficient	67%
1.2	Very Efficient	83%

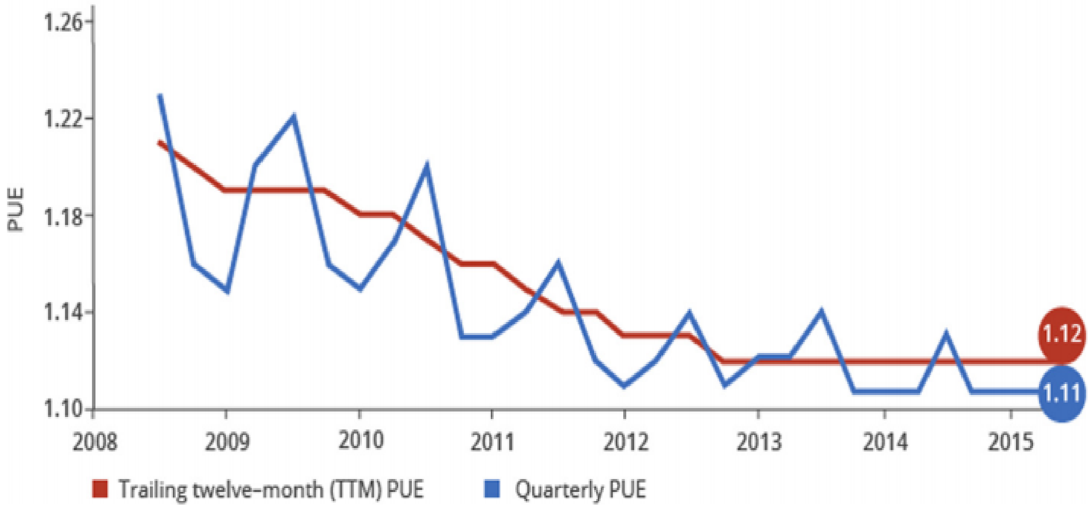


Figure 2.2: Continuous PUE Improvements (Average PUE)

2.1.5 Cost of a Data Center: Capital and Operational Expenses

When it comes to cost, according to J. Hamilton approximately 45% of data center costs are attributed to servers, including CPU, memory, and storage systems. Building a data center typically costs between $600 to $1,100 per gross square foot or $7 million to $12 million per megawatt of commissioned IT load. For a data center with 50,000 servers priced at $3000 each, the annual amortized cost is $52.5 million. Detailed cost breakdown shows that servers are 45% of the total expenditure, infrastructure 25%, power 15%, and networking 15% as major expenses. Operational staff costs are minimal almost less than 5% due to high automation **[21]**. The total installed cost for a high-density data center with 20,000 square feet of electrically active floor area is approximately \$100 million in the US. This includes 30% for initial IT equipment purchases and the rest for site infrastructure. The total electricity cost is \$2.46 million per year, with IT loads accounting for \$1.23 million, cooling \$0.80 million, and auxiliaries \$0.43 million on average. Infrastructure costs range from \$23,800 to \$26,000 per kW of useable UPS output, plus \$262 per square foot for electrically active floor area. Annual operating expenses include electricity costs (\$2.5 million), network fees (\$0.5 million), and other expenses such as management staff, maintenance, security, and property taxes totaling \$5.9 million **[22]**. However, giving exact figures for data center is very difficult and depends mostly on the type of data center and its sizes. The Uptime Institute's data center tiers shown in Table 2.5 describe different levels of reliability and redundancy for maintaining uninterrupted service. Tier I offers the most basic setup with just one path for power and cooling, so if something fails, it can cause downtime. Tier II improves on this by adding some backup systems, but still relies on a single path for power and cooling. Tier III enhances reliability with multiple active paths for power and cooling, allowing for maintenance without affecting operations. Tier IV provides the highest level of resilience with fully redundant systems and multiple active paths, ensuring that the data center stays operational even if there are failures **[23]**.

Table 2.5. Data Center Tier Levels and Their Characteristics

Tier	**Redundancy**	**Fault Tolerance**	**Cooling and Power**
Tier I	N	No fault tolerance	Single-path cooling and power
Tier II	N+1	Basic fault tolerance	Redundant cooling and power components
Tier III	N+1	Partial fault tolerance	Multiple active paths for cooling and power
Tier IV	2(N+1)	Full fault tolerance	Fully redundant cooling and power paths

When planning and maintaining a data center, understanding the comprehensive cost structure is crucial. These costs are generally divided into capital expenditures (CapEx) and operational expenditures (OpEx). Each category encompasses various components, from the land and building shell to the IT infrastructure and operational services. Below is a detailed exploration of these costs, followed by a summarized table.

2.1.5.1 Capital Expenditures (CapEx)

Capital expenditures refer to the upfront costs associated with the creation and setup of a data center. These include expenditures on land, construction, equipment, and initial setup services.

1. Land: Acquiring land is a significant initial cost, especially in prime locations with reliable power and connectivity. The cost varies widely depending on the location and size of the plot.
2. Building Shell: Constructing the physical structure of the data center, including foundations, walls, roofing, and basic infrastructure. This also includes costs for zoning, permits, and compliance with local building codes.
3. Power and Cooling Infrastructure: Installing electrical systems, backup generators, uninterruptible power supplies (UPS), and cooling systems. This ensures that the data center can maintain consistent operations and handle high power demands efficiently.
4. Networking Infrastructure: Setting up networking equipment such as switches, routers, and cabling. This also covers the costs of integrating with external telecommunications networks.
5. IT Equipment: Purchasing servers, storage devices, and other IT hardware necessary for data center operations. This is typically one of the most significant CapEx components.
6. Security Systems: Implementing physical and cybersecurity measures, including surveillance cameras, biometric access controls, fire suppression systems, and intrusion detection systems.
7. Professional Services: Costs associated with design, engineering, and consulting services required during the planning and construction phases. This also includes legal and financial services for securing funding and complying with regulations.

2.1.5.2 Operational Expenditures (OpEx)

Operational expenditures refer to the ongoing costs required to run and maintain the data center. These include energy costs, personnel, maintenance, and various recurring expenses.

1. Energy Costs: The largest component of OpEx, encompassing electricity for powering IT equipment, cooling systems, lighting, and other electrical needs. Efficient energy management is critical to controlling these costs.

2. Personnel: Salaries and benefits for staff involved in operations, maintenance, security, and management. This includes IT professionals, facility managers, security personnel, and administrative staff.

3. Maintenance and Repairs: Routine maintenance and repairs for IT equipment, power and cooling systems, and the building itself. This ensures continuous and reliable operation and extends the lifespan of the infrastructure.

4. Software and Licensing: Costs for software required to manage data center operations, including operating systems, virtualization software, monitoring tools, and security solutions. Licensing fees for proprietary software are also included.

5. Telecommunications: Expenses related to connectivity and bandwidth, including contracts with internet service providers and telecom companies to ensure robust and high-speed connections.

6. Insurance: Coverage for the data center infrastructure and operations against risks such as natural disasters, cyberattacks, and other potential hazards.

7. Miscellaneous Operating Costs: Additional operational costs such as taxes, regulatory compliance, and miscellaneous administrative expenses. Data center initial cost distribution percentages is shown in graph given in figure 2.3.

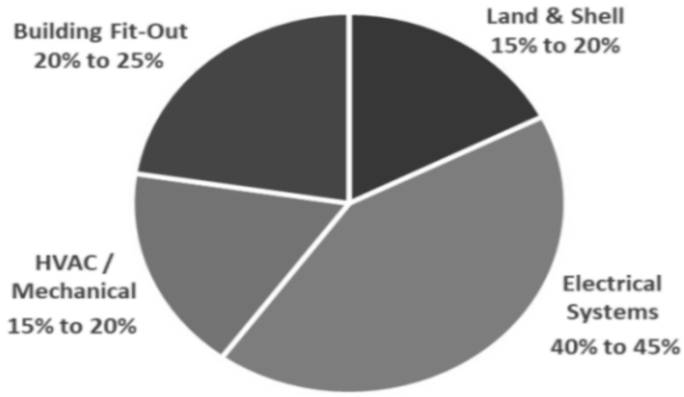


Figure 2.3: Data Center Capital Cost Breakdown

Chapter 3

Mathematical Models and Formulas

3.1 Data Center Equipment Power Models

3.1.1 Servers Power Models

The equation 3.1 from **[24]** illustrates how a server's power consumption varies with its utilization, which is essential for optimizing energy use in data centers. In this equation, P_i represents the power consumption of a server i at a given time, P_{idle} denotes the power consumed when the server is idle, and P_{peak} is the maximum power consumption when the server is fully utilized. The variable ui indicates the server's utilization at time i, scaled between 0 and 1. As the server's workload increases, the power consumption Pi increases linearly from P_{idle} to P_{peak}, highlighting the direct relationship between server utilization and power consumption. This model is crucial for designing strategies to manage energy efficiency in data centers, ensuring that energy use is proportional to the workload.

$$P_i = P_i^{idle} + (P_i^{peak} - P_i^{idle}) \times u_i \tag{3.1}$$

This equation reflects that even when a server is idle, it consumes a baseline amount of power P_{idle}, and as utilization u_i increases, power consumption P_i approaches its peak value P_i peak. So, the total power consumed by the server farm is the sum of all the powers consumed by each server unit.

$$P_{server} = \sum_{i=1}^{N} P_i \tag{3.2}$$

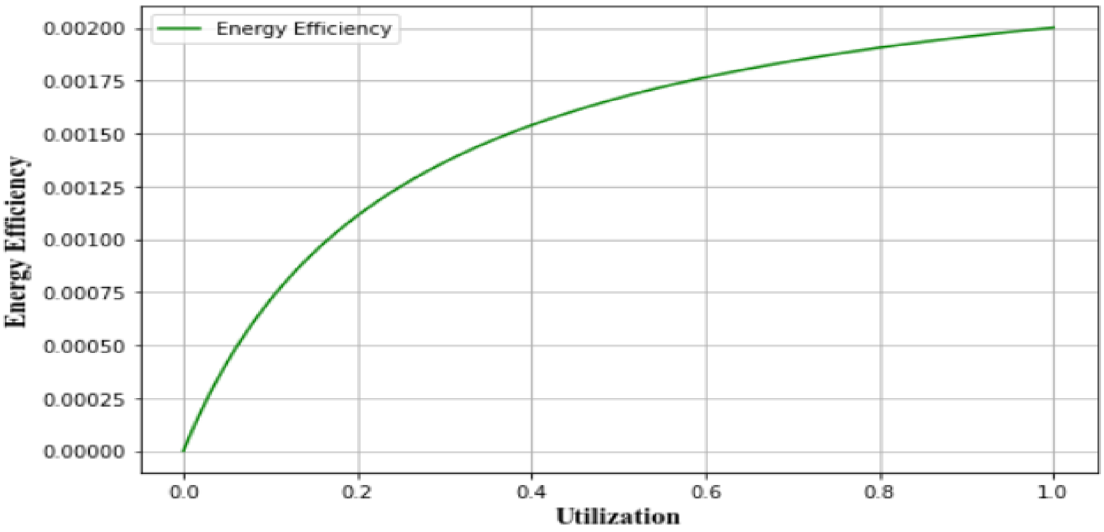


Figure 3.1: Servers Utilization vs Efficiency

If we modify this equation to find relation between server power consumption and energy efficiency which is defined as the ratio of useful work done to the energy consumed, represented as $E_{eff} = \frac{u_i}{P_i}$

Inserting 3.1 to it we derive 3.3 further.

$$E_{eff} = \frac{u_i}{P_{idle} + P_{peak} - P_{idle} \times u_i} \tag{3.3}$$

Energy efficiency E_{eff} inversely relates to the power consumption at a given utilization.

Higher u_i values improve E_{eff}, demonstrating better energy use as servers handle more workload. Figure 9 shows inverse relation between power P_i and efficiency E_{eff}.

3.1.2 Power Conditioning Units

The power conditioning and supply system of a data center is a complex infrastructure designed to ensure uninterrupted and high-quality electrical power supply for smooth functioning. The primary source of electrical power is the utility power supply, which provides the initial electricity for the data center but is subject to fluctuations. To address these challenges, the uninterruptible power supply (UPS) systems are employed, offering immediate backup power during outages and voltage fluctuations. These systems typically utilize batteries to maintain power until the generators are activated or utility power is restored.

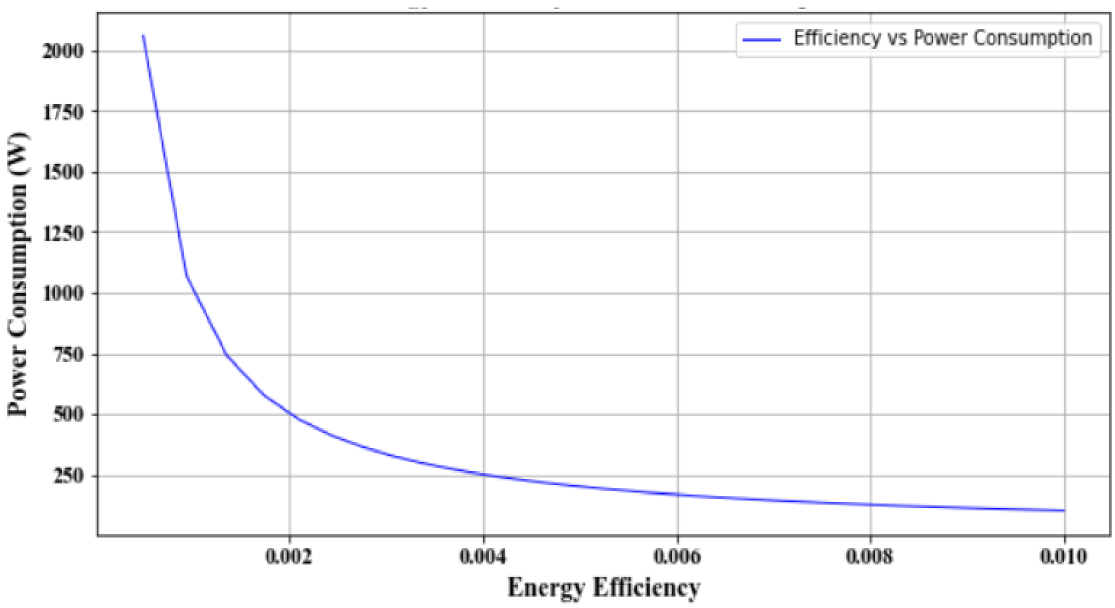


Figure 3.2: Servers Power Consumption vs Efficiency

The power distribution units (PDUs) distribute electricity from the UPS and generators to the various racks and equipment within the data center, and often come equipped with features such as load monitoring and power metering to ensure efficient and reliable power delivery. In the event of extended outages or utility failures, generators serve as backup power sources which are usually diesel-powered, and are integrated with automatic transfer switches to ensure a smooth transition of power. The automatic transfer switches (ATS) automatically switch the power source from the utility to the generator during a power failure, ensuring that critical systems remain operational without interruption. Together, these components form a power conditioning and supply system that supports the reliability and efficiency of the data center's operations. According to **[24]** the power consumed by a Power Distribution Unit (PDU) in a data center is directly related to the power consumption of the servers connected to it and its total power consumption includes both the power delivered to the servers and the inherent power losses in the PDU itself. These losses occur due to inefficiencies in power conversion and distribution, as well as heat dissipation and other operational factors. Equation 3.4 and 3.5 give power models for PDUs and UPS which both have idle power consumption of both components which shows that they consume energy even when idle.

$$P_{PDU} = P_{PDU}^{idle} + \lambda_{PDU} \left(\sum_{\text{servers}} P_{server_i} \right) \tag{3.4}$$

$$P_{PDU} = P_{UPS}^{idle} + \lambda_{PDU} \left(\sum_{\text{PDUs}} P_{PDU} \right) \tag{3.5}$$

Where, $lambda_{PDU}$ and $lambda_{UPS}$ are the loss coefficient for the PDU and UPS. This indicates that as the total power drawn by the servers increases, the power loss in the PDU increases exponentially. This model highlights the importance of minimizing server power consumption and consolidation to reduce overall power losses in PDUs.

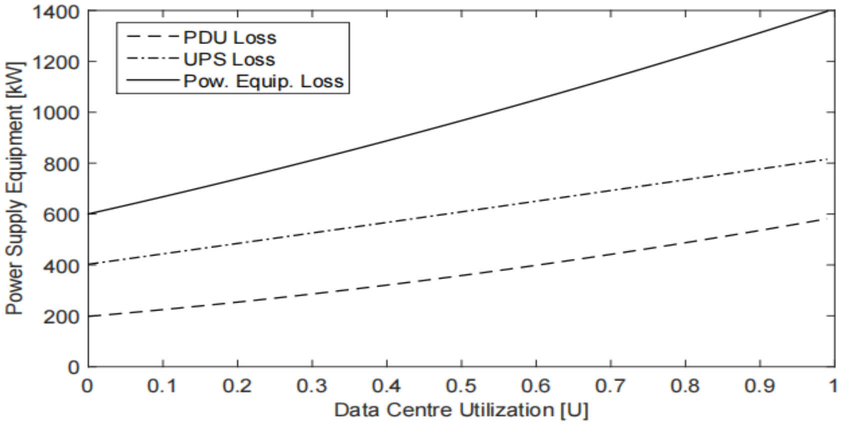


Figure 3.3: Power Loss of Power Conditioning Equipment vs Utilization

3.1.3 Cooling Equipment Power Models

Servers and other hardware generate a significant amount of heat during operation, similar to how other heavy machinery or electrical equipment does. Without proper cooling, this heat can build up, leading to hardware malfunctions, reduced performance, and even permanent damage. Effective cooling systems help dissipate this heat, maintaining an optimal temperature that ensures reliable performance and longevity of the equipment and for keeping the servers and other hardware running smoothly by preventing them from overheating, which can cause severe damage and expensive downtime. Moreover, efficient cooling can significantly reduce energy consumption, making data centers more environmentally friendly and cost-effective.

The power consumption of CRAC units can be modeled using thermodynamic principles considering the cooling capacity, coefficient of performance (COP), and the electrical power input. For CRAH units, models incorporate the energy consumption of chilled water pumps and fans. The models typically use parameters like water flow rate, temperature differential, and fan power laws to estimate consumption. The most relevant power models for data center cooling are given in **[25]** in which $P_{cooling}$ as in 3.6 is the overall summation of all other powers consumed by different cooling equipment for establishing the cooling infrastructure and fulfilling the requirements of data center.

$$P_{cooling,t} = P_{Fan,t} + P_{Blower,t} + P_{CWP,t} + P_{Chiller,t} + P_{CTP,t} + P_{CTF,t} \tag{3.6}$$

where, $P_{Fan,t}$ represents the power consumed by the server fans at time t. This is calculated based on the rate at which the fan moves air and the difference in the air's enthalpy as it passes through the fan whereas $P_{Blower,t}$ is determined by the mass flow rate of the air moved by the blower and the change in enthalpy of the air as it passes through the CRAC unit. $P_{CWP,t}$ denotes the power consumption of the chilled water pump (CWP) which is calculated based on the mass flow rate of the chilled water, the pressure drop across the system, and the efficiency of the pump while $P_{CTP,t}$ and $P_{CTF,t}$ signifies the power required for the cooling tower pump (CTP) and the power consumption of the cooling tower fan (CTF) respectively. $P_{CTP,t}$ is determined by the mass flow rate of the hot water, the pressure drop in the system, and the pump's efficiency at time t. $P_{Chiller,t}$ represents the power used by the refrigeration chiller and involves the work done by the compressor in the chiller system to compress the refrigerant gas, which is influenced by the mass flow rate of the refrigerant and the enthalpy changes in the system. Another model for power consumption of CRAC units is discussed in **[26]** as a function of cooling load Q which is the inlet heat flowing into the CRAC and coefficient of performance of CRAC units as shown in 3.7.

$$P_{CRAC} = \frac{Q_{cool}}{\text{CoP}} \tag{3.7}$$

As the Coefficient of Performance (CoP) increases, the CRAC unit needs less cooling power because a higher CoP means the unit is better at removing heat with the same amount of energy. For cooling equipment that has higher CoP values, it reduces the power consumed for cooling data centers significantly and hence has a direct impact on overall energy savings and reduced operational costs. This equation shows relation between efficiency and power consumption by CRAC unit through parameter CoP. Another metric of calculating efficiency of CRAC units is the Energy Efficiency Ratio EER which is a function of cooling output in British Thermal Units per hour and electrical power input in Watts as expressed in 3.8.

$$\text{EER} = \frac{\text{Cooling Output (BTU/hr)}}{\text{Electrical Power Input (W)}} \tag{3.8}$$

Figure 3.4 derived from **[27]** shows EER values during different outside temperatures which shows that the efficiency of cooling systems in data centers is significantly influenced by the ambient temperature.

Cooler outside temperatures enhance the Energy Efficiency Ratio (EER) of chiller plants, meaning they operate more efficiently and consume less energy. Therefore, cooler external conditions can lead to substantial energy savings and reduced operational costs for data centers. For instance, As discussed in **[24]**, during a winter week the chiller's contribution to energy consumption drops to 16%, decreasing from 4.02610^8 kWh to 2.588410^8 kWh compared to summer. Additionally, the data center's total energy consumption in summer is 7.88% higher than in winter, totaling 2.030610^9 kWh, with varying power consumption profiles depicted for different temperatures.

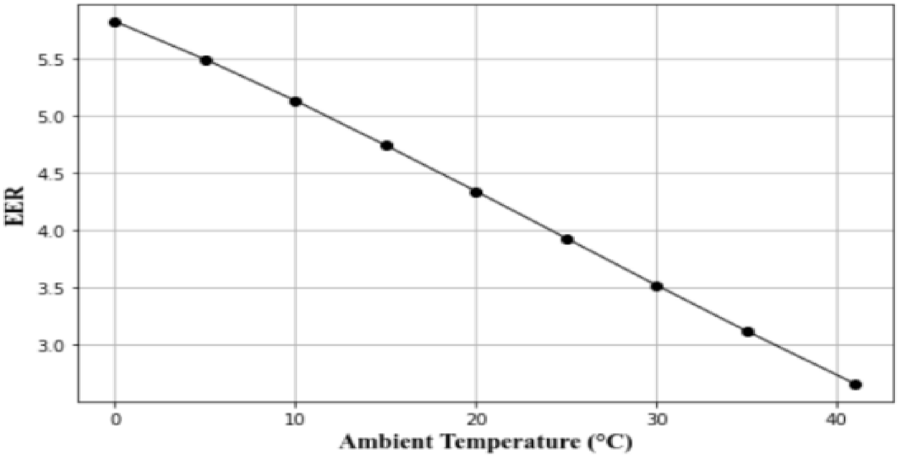


Figure 3.4: Outside Temperature Impact on CRAC EER

3.2 Financial Analysis Methods

Investing in green data centers often involves significant upfront costs, which can be a substantial negative cash flow for organizations making the transition. This initial expense is associated with the installation of energy-efficient technologies and infrastructure upgrades, including advanced cooling systems and renewable energy sources. However, despite the high initial investment, the long-term benefits can outweigh these early financial setbacks. Over time, the reduced energy consumption leads to considerable savings on utility bills, contributing to a more favorable financial outlook. Additionally, operational costs are lowered due to increased efficiency and decreased maintenance needs. As a result, these savings can gradually offset the initial expenditure, ultimately yielding a positive return on investment (ROI). By focusing on long-term financial gains rather than short-term costs, companies can achieve a more sustainable and profitable operational model. Thus, while the transition to green data centers may be costly at the outset, the overall financial benefits and environmental advantages make it a worthwhile investment. This section explores various cost models for green data centers and discusses several cash flow analysis tools.

3.2.1 Total Cost of Ownership TCO

Total Cost of Ownership (TCO) of a data center is the complete cost to build and run it over its lifetime. This includes the initial costs of constructing the facility, buying equipment, ongoing expenses like electricity, cooling, and maintenance. Total cost of ownership (TCO) is a crucial factor for data center stakeholders when determining the level of energy efficiency in their designs. TCO includes both capital expenditures (CapEx) and operational expenditures (OpEx). Unfortunately, CapEx often overshadows OpEx in TCO calculations. The misconception that energy-efficient data centers lead to higher TCO is misleading because OpEx is not always accurately accounted for. Table 5.1 **[28]** shows different capital expenses known as CapEx which includes the price of buying the hardware and software and the operational expenses OpEx for TCO of a data center. TCO can be calculated as given in equation below.

$$TCO_T = \text{Initia Costs } + \left(\sum_{t=1}^{T} \text{Direct Costs} + \text{Indirect Costs} \right) \tag{3.9}$$

Furthermore, the cost of the physical space that houses the data center constitutes a significant portion of the TCO, influenced by location-specific factors such as real estate prices, cooler locations and local utility rates. Throughout the life of a data center, the regular need to upgrade and replace outdated or failing equipment adds notable expenses. Additionally, adhering to regulatory standards and requirements incurs considerable costs for audits and certifications. Taxes and compliance costs, which vary by location, also influence the total financial outlook. These ongoing financial demands highlight the challenge of balancing efficiency, security, and regulatory adherence in a continually evolving tech landscape.

Table 3.1. Data Center TCO

Capital Costs	**Operational Costs**
Hardware Price (IT devices, power distribution, cooling equipment)	Power Consumption (electricity for servers, networking equipment, cooling)
Basic Installation, Design & Engineering Costs	Personnel Cost (salaries)
Land Cost	Maintenance and Repairs Costs
Building Cost	Amortization Costs
Project Management / Facility Engineering Cost	IT Software License Cost

3.2.2 Financial Analysis Metrics

Financial metrics are essential tools for assessing a company's financial performance and overall health. They include key figures like profit margins, which show how much profit a company makes from its revenue, and return on investment (ROI), which helps evaluate the effectiveness of investments. By analyzing these metrics, businesses can make informed decisions, identify strengths and weaknesses, and plan for future financial strategies. Data center financial analysis involves key metrics such as Payback Period, Return on Investment (ROI), Net Present Value (NPV), and Rate of Return to assess the viability and profitability of investments.

3.2.2.1 Payback Period

The Payback Period measures how long it takes for an investment to recoup its initial costs,with a shorter payback period suggesting a quicker return on investment. It is a ratio of initial investment to the annual cash inflow 3.10. It will be the analysis tool most widely used to assess the green investments made in data centers which incur a huge upfront cost, but through savings on energy and operational expenses they have a shorter payback period and thus support the hypothesis that green investment is good for businesses financial stability and economic sustainability.

$$\text{Payback Period} = \frac{\text{Initial Investment}}{\text{Annual Cash Inflows}} \tag{3.10}$$

3.2.2.2 Return on Investment

Return on Investment (ROI) calculated by 3.11 measures the profitability of an investment by comparing the profit made to the amount of money invested. A high ROI indicates that the investment has generated significant profit relative to its cost, suggesting it is a successful and efficient use of resources. It is calculated as a percentage ratio of net profit generated by the investment to the cost of initial investment.

$$\text{ROI} = \frac{\text{Net Profit}}{\text{Cost of Investment}} \times 100 \tag{3.11}$$

3.2.2.3 Net Present Value

Net Present Value (NPV) measures the value of an investment by considering the time value of money. It determines whether an investment is profitable by calculating the difference between the present value of cash inflows and the initial investment cost as shown in 3.12. A positive NPV indicates that the investment is expected to generate more money than it costs, making it a good investment.

$$\text{NPV} = \sum \frac{\text{Cash Inflows}_t}{(1 + \text{Discount Rate})^t} - \text{Initial Investment} \tag{3.12}$$

where "t" represents each time period. Understanding these metrics helps stakeholders make informed decisions by evaluating the time required to recover investments, the profitability ratio, the overall value considering future cash flows, and the percentage return expected from the data center investment.

Chapter 4

Data Center Certifications and Frameworks

Data centers should pursue certifications like LEED, ISO 50001, ENERGY STAR, DCMM, DCeP, EPEAT, the EU Code of Conduct, and TIA-942 to enhance their operational efficiency, sustainability, and reliability. LEED certification focuses on energy and environmental design, promoting sustainability in construction and operation. ISO 50001 provides a framework for continuous energy management improvements, helping data centers reduce energy consumption and greenhouse gas emissions. ENERGY STAR certification recognizes top-performing data centers for their energy efficiency, leading to lower operational costs. The Green Grid DCMM offers a maturity model for evaluating and improving efficiency and sustainability practices across data center operations. The US DOE's DCeP program provides resources and best practices to achieve significant energy savings. EPEAT certification ensures that data centers use environmentally friendly electronic products, supporting sustainability goals. The EU Code of Conduct for Data Centers encourages best practices in energy efficiency and environmental impact reduction within the EU. TIA-942 sets standards for data center design and operation, ensuring high availability, security, and performance. By obtaining these certifications, data centers can demonstrate their commitment to sustainability, enhance their reputation, attract environmentally conscious clients, and achieve long-term cost savings through improved energy efficiency and operational excellence.

4.1 Leadership in Energy and Environmental Design (LEED) Certification

The LEED Certification, established by the U.S. Green Building Council, assesses buildings on sustainability, energy efficiency, and environmental impact. Data centers aiming for LEED certification must optimize energy use, reduce water consumption, and improve indoor environmental quality. LEED evaluates performance across various categories, including site sustainability, water efficiency, energy and atmosphere, materials and resources, and indoor environmental quality. Achieving LEED certification can significantly enhance a data center's operational efficiency and reduce its carbon footprint. This certification is recognized globally and often serves as a benchmark for environmental stewardship in the construction and operation of data centers.

4.2 ISO 50001

ISO 50001 is an international standard for energy management systems, helping organizations develop a policy for more efficient energy use and establish targets to meet energy goals. Data centers adopting ISO 50001 implement systematic energy management practices, enhancing energy performance and reducing greenhouse gas emissions. The standard promotes continuous improvement in energy efficiency, integrating energy management into everyday organizational practices. Certification requires rigorous documentation and regular audits to ensure compliance and effectiveness. By achieving ISO 50001, data centers demonstrate a strong commitment to sustainable energy management and operational excellence.

4.3 ENERGY STAR Certification for Data Centers

The U.S. Environmental Protection Agency's ENERGY STAR Certification program honors data centers with exceptional energy efficiency performance. To earn this certification, data centers must meet strict energy performance standards, typically within the top 25 percent of facilities nationwide. The certification process involves benchmarking energy performance using ENERGY STAR's Portfolio Manager tool and achieving a score of 75 or higher. Certified data centers benefit from reduced operational costs, enhanced environmental performance, and increased marketability. This certification is a mark of energy efficiency excellence, encouraging continuous improvement and innovation in energy management practices.

4.4 Green Grid Data Center Maturity Model (DCMM)

The Green Grid Data Center Maturity Model (DCMM) provides a comprehensive framework for evaluating and improving data center efficiency and sustainability. It outlines five levels of maturity across several key areas, including power, cooling, compute, storage, and network infrastructure. Data centers use the DCMM to assess their current performance, identify areas for improvement, and implement best practices to enhance energy efficiency and sustainability. The model encourages continuous improvement and helps data centers achieve higher levels of operational efficiency and environmental stewardship. By following the DCMM, data centers can systematically reduce their energy consumption and carbon footprint.

4.5 US Department of Energy (DOE) Data Center Energy Efficiency Program (DCeP)

The US Department of Energy's Data Center Energy Efficiency Program (DCeP) provides resources and best practices to improve the energy efficiency of data centers. The program focuses on optimizing energy use in IT equipment, cooling systems, and power distribution. DCeP offers tools, training, and technical assistance to help data centers implement energy-saving measures and achieve significant cost savings. The program promotes collaboration among industry stakeholders to drive innovation and share successful strategies for energy management. By participating in DCeP, data centers can enhance their energy efficiency, reduce operational costs, and contribute to national energy sustainability goals.

4.6 Electronic Product Environmental Assessment Tool (EPEAT) Certification

The Electronic Product Environmental Assessment Tool (EPEAT) Certification evaluates electronic products, including data center equipment, based on their environmental impact. EPEAT criteria cover various aspects such as energy efficiency, recyclability, reduction of toxic materials, and product longevity. Data centers that procure EPEAT-certified equipment ensure that their operations are more sustainable and environmentally friendly. The certification promotes the use of greener technologies and supports the circular economy by encouraging the design and manufacture of more sustainable electronic products. EPEAT certification helps data centers meet environmental regulations and enhance their corporate social responsibility.

4.7 EU Code of Conduct for Data Centers

The EU Code of Conduct for Data Centers provides a framework for improving energy efficiency and reducing the environmental impact of data centers within the European

Union. It outlines best practices for energy management, including optimizing cooling and power systems, enhancing IT equipment efficiency, and adopting renewable energy sources. Data centers that adhere to the Code of Conduct commit to continuous improvement and regular reporting on energy performance. The initiative fosters collaboration and knowledge sharing among data center operators, vendors, and policymakers. By following the Code of Conduct, data centers can achieve significant energy savings and support the EU's sustainability goals.

4.8 TIA-942

TIA-942 is a standard developed by the Telecommunications Industry Association that outlines requirements for data center design, construction, and operation. It addresses critical aspects such as site location, architectural considerations, electrical and mechanical systems, fire safety, and network connectivity. The standard defines four levels of data center reliability, from basic infrastructure to fault-tolerant systems, ensuring high availability and robust performance. Compliance with TIA-942 helps data centers achieve optimal efficiency, security, and resilience. This standard is widely adopted and respected in the industry, providing a benchmark for data center quality and operational excellence.

4.9 Data Center Energy Efficiency Framework (DCEEF) by the Green Data Center Alliance

The Data Center Energy Efficiency Framework (DCEEF) by the Green Data Center Alliance provides a comprehensive approach to enhancing energy efficiency in data centers. This framework focuses on practical strategies for reducing energy consumption while maintaining optimal performance. Key areas of improvement include optimizing cooling systems, improving power usage effectiveness (PUE), and managing server operations more efficiently. Implementing DCEEF helps data centers identify and address inefficiencies, leading to more sustainable and cost-effective operations. The framework also emphasizes the importance of regular monitoring and assessment to ensure continuous improvement. By following DCEEF guidelines, data centers can significantly lower their energy costs and reduce their environmental impact. This approach not only supports operational efficiency but also aligns with broader sustainability goals. Overall, DCEEF serves as a valuable resource for data centers striving to achieve higher standards of energy efficiency and environmental responsibility.

4.10 Guidelines for Energy-Efficient Datacenters by The Green Grid

The Green Grid's Guidelines for Energy-Efficient Datacenters offer a detailed roadmap for improving energy efficiency in data center operations. These guidelines provide actionable strategies for optimizing various aspects of data center infrastructure, including cooling systems, power distribution, and server utilization. By implementing these recommendations, data centers can achieve significant reductions in energy consumption and operational costs. The guidelines advocate for the use of advanced cooling technologies and energy-efficient hardware to minimize waste and enhance performance. Additionally, The Green Grid emphasizes the importance of leveraging renewable energy sources to further reduce the environmental footprint of data centers. Continuous monitoring and evaluation are critical components of these guidelines, ensuring that data centers maintain high efficiency levels over time. Adhering to The Green Grid's guidelines not only boosts energy efficiency but also promotes sustainable practices within the industry. These comprehensive strategies help data centers become more resilient and environmentally friendly, benefiting both the business and the planet.

Chapter 5

ROI Analysis Of Green Infrastructure Through Energy Savings

In this section we show that how the initial green investment which is a negative cash outflow is beneficial for data centers environmental as well as economic sustainability as this investment reduces the carbon footprint, improves PUE while also reducing the cash flowing out to pay for energy bills the major contributors of which are servers, cooling infrastructure and power conditioning units. The analysis is done on several case studies and data on costs of different sustainable options and their potential energy and operational savings collected from different sites such as Green Grid, Energy Star, LAZARD, ASHRAE, and Schneider Electric. ROI and payback period analysis is done on the collected data to help advisors and investors in decision making on investment in green infrastructure and equipment in cloud data centers which incur high upfront costs but have greater ROI compared to traditional cloud solutions and thus after the payback period is over, which is achieved through cumulative cost of generated revenue, reduced maintenance costs and energy billing savings combined, the savings start benefiting the data centers and start generating actual revenue **[29]** .

5.1 Energy Savings and ROI Analysis of Green Infrastructure

5.1.1 Savings Through Energy Efficient Servers

Energy-efficient servers are designed to minimize power consumption while maintaining high performance and reliability. These servers utilize various technologies and strategies to optimize energy usage, making them ideal for reducing operational costs, energy billing and environmental impact in data centers**[30]**. These servers have ability to adjust power usage based on workload. This includes features like dynamic voltage and frequency scaling (DVFS), which adjusts the processor's power and performance according to the workload demand and deliver optimized performance per watt which is metric of server efficiency measured in FLOPS/Watt which is improving through the years Figure 5.1 **[31]**, meaning they provide higher computational power for each watt of energy consumed **[32]**.

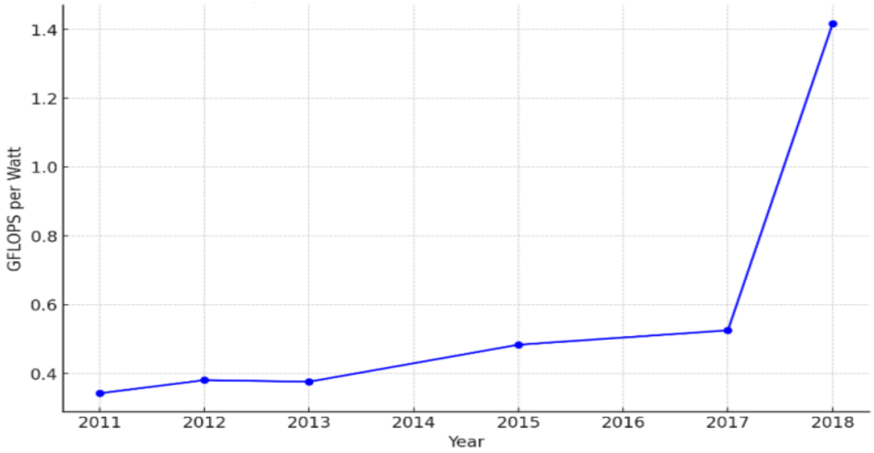


Figure 5.1: Performance per Watt of Intel i7 Core Processors Over the Years

Energy-efficient servers often have higher thermal limits, allowing them to operate effectively at higher temperatures compared to traditional servers. The most interesting findings on cost savings through efficient servers is found in **[33]**. The paper provides detailed case studies of three organizations to demonstrate the financial benefits of hardware refreshes in small data centers. The study focuses on three different configurations for upgrading servers. In original configuration, servers are upgraded to the latest version without changing the number of servers while in equal computation configuration the computational resources are maintained, but the number of servers is reduced due to improved processor performance and lastly in high utilization configuration the number of servers is further reduced to increase utilization from 40% to 50%. The intersection point of cumulative costs shows that ROI can be achieved in less than two years for the first two organizations and less than three years for the third which is achieved by significant reductions in energy costs are achieved through hardware refresh and optimization of supporting equipment.

Jobs are scheduled on servers to efficiently manage computational resources and ensure smooth operation. This scheduling assigns tasks to available servers based on their current load and the resources required by the tasks. As these jobs run, they utilize the server's CPU cycles, memory, and storage, increasing the server's utilization. There is a direct relationship between utilization and power consumption: as the server's utilization rises with more jobs, its power consumption also increases. Figure 5.2 shows Dell PowerEdge 2970 and R710 utilization vs power consumption comparison. This means that even when servers are idle, they consume a baseline amount of power, but this consumption grows significantly with increased computational load. Efficient scheduling aims to balance the load across servers to optimize their performance and minimize unnecessary power consumption, thus enhancing overall energy efficiency. Variables that influence how much energy servers consume are input/output, CPU utilization percentage, server hardware, software, and the quantity of storage access needed for a particular task.

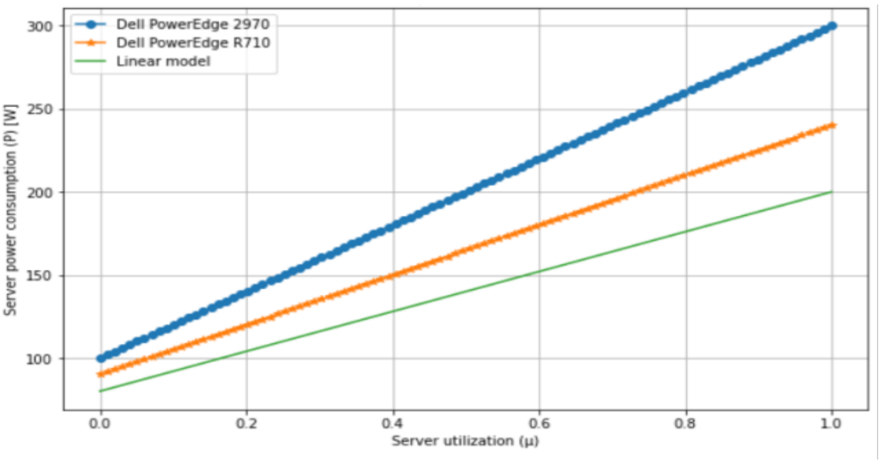


Figure 5.2: Server Utilization vs Power Consumption Relation

5.1.1.1 Cost and Power Analysis Old vs Energy Efficient Server

To conduct the payback period analysis for the data center, we followed a case study involving the replacement of traditional HP Proliant DL360 G5 servers with energy-efficient G6 servers which were ENERGY STAR certified servers. The following steps were taken:

1. **Initial Investment Costs:**
   The cost of a traditional G5 server was assumed to be approximately $1,000 each. G6 Servers: The cost of an energy-efficient G6 server was assumed to be approximately $1,500 each. Total Cost for G5 Servers: For 200 servers, the total cost was calculated as 200 * $1,000 = $200,000. Total Cost for G6 Servers: For 200 servers, the total cost was calculated as 200 * $1,500 = $300,000.

2. **Energy Consumption Differences:**
   Servers certified by ENERGY STAR offer that extra computational power utilizing about 30% less energy, based on estimates from the EPA. So, if each G5

server consumes 450 watts and each G6 server consumes 315 watts per hour on a 40% utilization. Also they reduce energy utilized when servers are idle by 10%.

3. **Methodology:**
   The HP ProLiant DL360 G6, which was installed with a new operating system (OS) installation (Windows Server 2008 R2), was the new ENERGY STAR-qualified server. This was contrasted with an older, non-ENERGY STAR rated HP ProLiant DL360 G5 running Windows Server 2003 Service Pack 2. Additionally, a brand-new OS installation and the stock configuration were applied to the G5.

4. **Workloads:**
   Three workloads were selected for testing: an industry-standard power and performance workload (run as a baseline test), Web Fundamentals, File Server Capacity Tool (FSCT)

5.1.1.2 Energy Consumption and Performance/Watt Difference

The G6 servers demonstrate significantly higher power efficiency than the G5 servers at every load level, with efficiency improvements ranging from 221% to 332%. For example, at 100% load, the G6 server has a power efficiency of 1369 performance/watt, whereas the G5 server has only 427 performance/watt, representing a 221% improvement. Additionally, the G6 servers consistently consume less power than the G5 servers, with differences in power consumption ranging from 11% to 54%. At lower load levels, this difference becomes even more pronounced, highlighting the superior energy efficiency of the G6 servers. In terms of raw performance, the G6 servers consistently outperform the G5 servers, delivering higher transactions per second across all load levels. This improvement in performance, coupled with reduced power consumption, underscores the G6 servers' ability to provide better performance per watt. The table also indicates that the power consumption of the G6 servers is significantly lower at idle, consuming only 119 watts compared to the G5 server's 256 watts, which highlights the G6 servers' efficiency even when not under load. Overall, the G6 servers offer substantial improvements in both performance and energy efficiency, making them a more cost-effective and environmentally friendly option. The detailed comparison is given in Table 8.

$$\text{Performance per WATT} = \frac{\text{Computing Performance}}{\text{Power Consumption}} \tag{5.1}$$

Table 5.1. G5 vs G6 Servers Power Consumption Difference at Different Loads

Load Level	**G5 Avg. Power (W)**	**G5 Power Efficiency**	**G6 Avg. Power (W)**	**G6 Power Efficiency**	**Difference in Power Consumed**	**Difference in Power Efficiency**
100%	346	427	307	1369	11%	221%
50%	291	253	220	959	24%	279%
10%	262	60	167	250	36%	332%

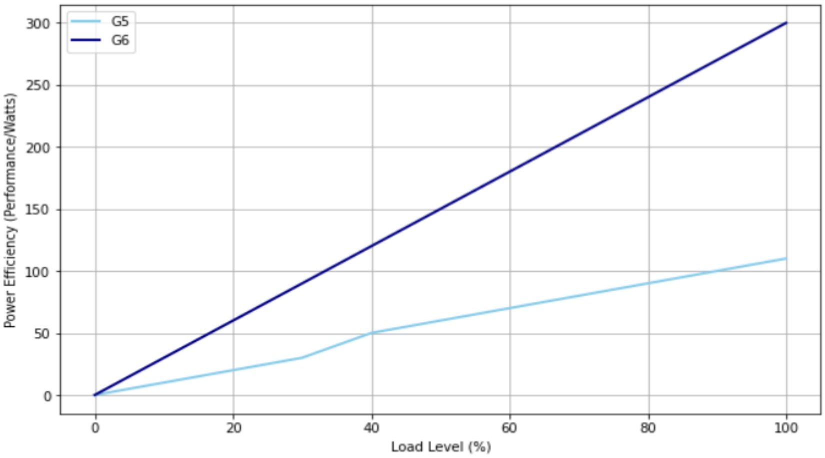


Figure 5.3: G5 vs G6 Server Performance/Watt Difference on Web Fundamentals Workload

5.1.1.3 Results

On average, the G6 server uses 26% less power than the G5 server across various workloads and delivers performance-to-power ratios that are 271% higher on average compared to the G5. After the initial investment cost comparison for scenario in which 200 old G5 servers are replaced with 200 new G6 servers, it was concluded that cost savings on energy and higher performance provided by G6 servers the payback period for this initial investment is approximately 5 years which is calculated through data collected on average cost of both servers gathered through archives of HP website, the cost of energy saved based on assumption that it costs $0.12 per kWh. The energy savings from a single ENERGY STAR-certified server vary from $60 (at 50% usage) to $120 (at idle) yearly, or $240–$480 over the usable life of a server (5 years), assuming the typical US business rate for electricity is 10 cents per kilowatt hour (kWh).ENERGY

STAR certified servers not only use less energy but also significantly lower data center cooling demands. According to basic guidelines, a server can also save one to two watts of avoided cooling for every watt it saves. This results in savings of $480 to $1,440 total throughout the course of a server's usable lifetime.

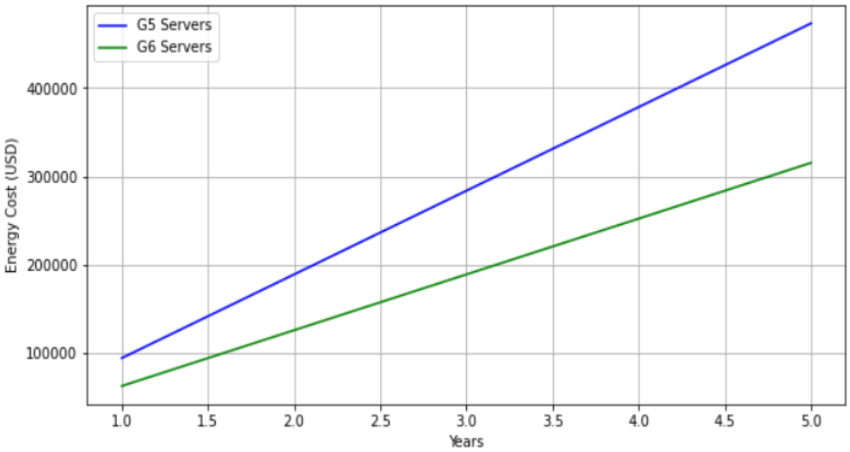


Figure 5.4: G5 vs G6 Servers Consumed Energy Cost Over the Years

By reducing power consumption, the G6 servers not only save on direct electricity expenses but also lower cooling requirements, which further contributes to energy savings. These energy savings are sufficient to offset the initial investment within a few years, demonstrating a clear return on investment (ROI). Moreover, implementing server virtualization and consolidation can further enhance these savings. Virtualization allows multiple virtual servers to run on a single physical server, maximizing resource utilization and reducing the number of physical servers needed. This reduction in the number of active servers directly decreases overall power consumption and cooling needs. Consolidation also optimizes server workloads, ensuring that servers operate closer to their optimal performance levels, which improves their power efficiency. However, effect of these energy efficient resource management techniques is not assessed here.

Table 5.2. Cost and Savings Analysis for Servers

Item	**G5 Servers**	**G6 Servers**	**Total Savings**
Total Cost of 200 Servers ($)	$200,000	$260,000	-
Annual Energy Consumption per Server (kWh)	3,942	2,228	-
Total Annual Energy Consumption (kWh)	788,400	445,600	262,800 (reduction)
Annual Energy Cost ($ (0.12$))	$94,608	$53,372	$41,236
Cooling Energy Savings (kWh)	-	-	62,560
Cooling Energy Savings Cost ($)	-	-	$7,507.20
Total Annual Savings ($)	-	-	$48,743.20
Payback Period (Years)	-	-	Approx 5 years

The graph in Figure 16 show that after 2.6 years the initial investment cost for 100 servers is paid back just by cost saved through energy savings. There are other factors that can reduce this payback period even more if taken into account for example the improved performance to watt ratio of G6 servers which would deliver more GFLOPS per watt and would take less time to complete the workload and would save on cooling costs due to higher thermal limit and efficient built in cooling mechanisms.

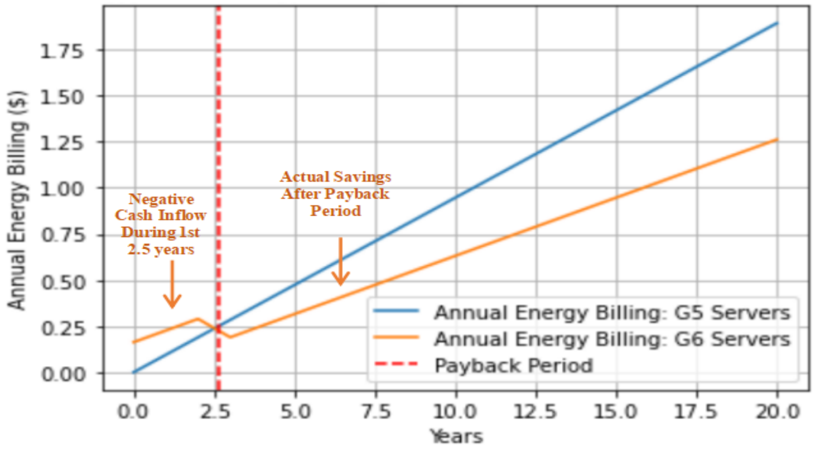


Figure 5.5: HP G5 vs G6 Servers Energy Billing

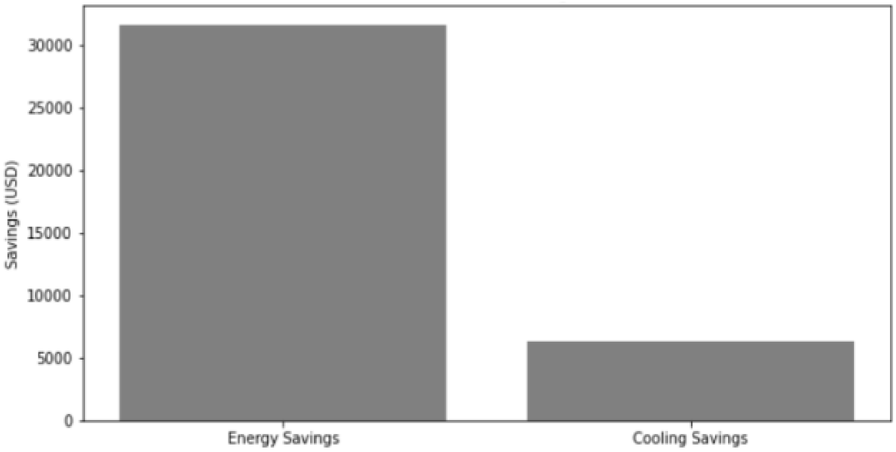


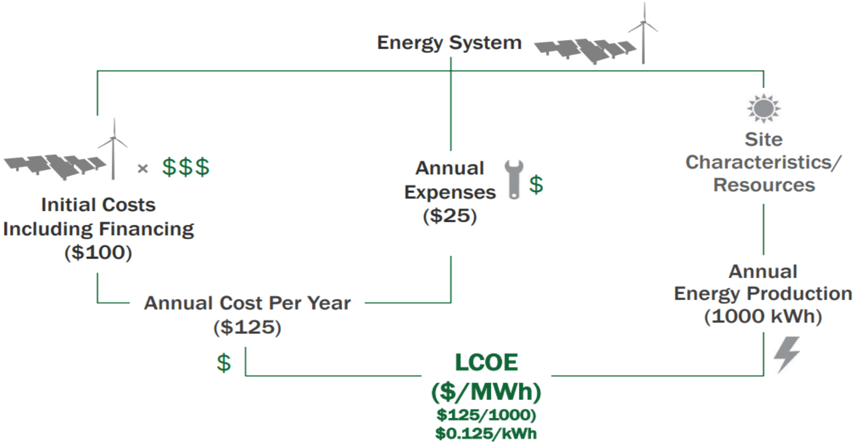
Figure 5.6: HP Proliant G5 vs G6 Servers Energy and Cooling Savings

5.1.1.4 Conclusion

Energy conservation at server farm reduces demand on the electrical system, which in turn leads to less electricity being produced and, consequently, less pollution. Based on the aforementioned assumptions, our analysis indicates that a single ENERGY STAR-qualified server saves enough electricity to prevent close to $\frac{1}{2}$ to 1 ton of carbon dioxide emissions. One to three tons of carbon dioxide are avoided overall when cooling savings are taken into consideration. As far as financial benefits are concerned, we have shown above that investing in energy efficient servers is a good choice for large scale data centers but for smaller data centers, investing in new servers can be expensive and payback period analysis shows that savings produced through efficient servers are paid back in 5 year period which is the average life time of a server before it needs upgrade.

5.1.2 Savings Through Renewable Energy Sources

Energy costs are the biggest expense in a data center and soon might exceed the hardware cost **[35]**. Investing in renewable energy sources for running the data centers is the most cost effective and best solution for lowering the PUE of data centers as well as saving on cost of billing. Energy produced through coal and other brown sources are most expensive and most harmful for environment yet most widely used. If a data center invests in renewable energy sources, it can lower its PUE and can provide a huge return on investment, can generate more revenue and attract more customers who are becoming more aware of climate change. The cost of brown energy sources is rising rapidly making it a difficult choice for organizations to run their infrastructures. Also the brown energy sources are highly influenced by different factors such as rising fuel prices, carbon taxes, and increasing capital costs. While trend suggests that governments around the world are providing subsidies on green energy sources, capital costs are decreasing at rapid scale and are not influenced by fuel costs. By 2022, the use of renewable energy around the world since 2000 has expected to have reduced fuel costs in the electrical industry alone by USD521 **[36]**. Load balancing can be used to direct traffic to data centers that utilize more renewable energy.



s

Figure 5.7: LCOE of Energy Systems

The Levelized Cost of Energy (LCOE) **[37]** is a metric used to compare the cost-effectiveness of different energy generation technologies. It represents the average total cost to build and operate a power-generating asset per unit of electricity produced over the asset's lifetime expressed in terms of currency per megawatt-hour (e.g., \$/MWh) as shown in eq 5.2.

$$\text{LCOE} = \frac{\sum_{t=1}^{T} \frac{I_t+O_t+F_t}{(1+r)^t}}{\sum_{t=1}^{T} \frac{E_t}{(1+r)^t}} \tag{5.2}$$

where it represents the investment expenditures for building the energy facility, Ot denotes the operations and maintenance costs, Ft is the fuel expenditure, Et indicates the amount of electricity generated, r is the discount rate reflecting the time value of money, and T is the project's expected lifetime. Green energy sources, such as solar and wind power, benefit significantly from government subsidies which reduce their overall costs and enhance their competitiveness. Subsidies in the form of Investment Tax Credits (ITC) and Production Tax Credits (PTC) play a crucial role in lowering the Levelized Cost of Energy (LCOE) for renewable technologies. For instance, the LCOE for solar photovoltaic (PV) utility-scale projects significantly decreases with subsidies, thus making them more attractive to investors and consumers . Fuel costs are a primary factor influencing the cost structure of brown energy sources, such as coal and natural gas plants. These costs are highly variable and depend on market prices for coal and natural gas, which can fluctuate significantly. For coal-fired plants, fuel costs can range significantly, and this variability can impact the overall cost-effectiveness and operational costs of these plants. Carbon pricing introduces an additional cost element for brown energy sources, aimed at internalizing the environmental costs associated with carbon emissions. The analysis shows that carbon pricing can significantly increase the LCOE for coal and natural gas plants, making them less competitive compared to renewable energy sources. For example, implementing a carbon price of $20 to $40 per ton of CO2 by data center can elevate the LCOE for conventional coal and gas plants above the levels of renewable sources like onshore wind and utility-scale solar **[38]**. This pricing mechanism is intended to incentivize reductions in carbon emissions and encourage a shift towards cleaner energy alternatives. These factors and analysis suggest that investing in green energy sources is far more beneficial financially for data centers.

In **[38]**, the unsubsidized analysis of LCOE of renewables such as solar PV, offshore wind, geothermal shows that their upfront cost is higher compared to the brown sources such as coal, nuclear and gas peaking but sensitivity to tax subsidies, sensitivity to fuel prices fluctuations and sensitivity to carbon pricing brings a huge difference. shows a significant reduction in capital costs for renewable energy sources like wind and solar photovoltaic (PV), making them increasingly cost-competitive with traditional energy sources like gas and coal. This trend highlights the advancements in technology and increased competition driving down costs in the renewable energy sector. Based on the cost data comparisons provided in **[38]**, we do payback period analysis for a 1MW data center that invests in renewables. This analysis includes initial capital investment costs that includes costs associated with deploying green energy infrastructure (e.g., solar panels, wind turbines), operational costs which include annual operational costs for both green and brown energy sources, Cost per kWh for green and brown energy, government subsidies and incentives, carbon taxes, and fuel costs. After the calculations it shows that all these expenses combined have a pay back period of approximately 4.9

years through energy cost comparisons.

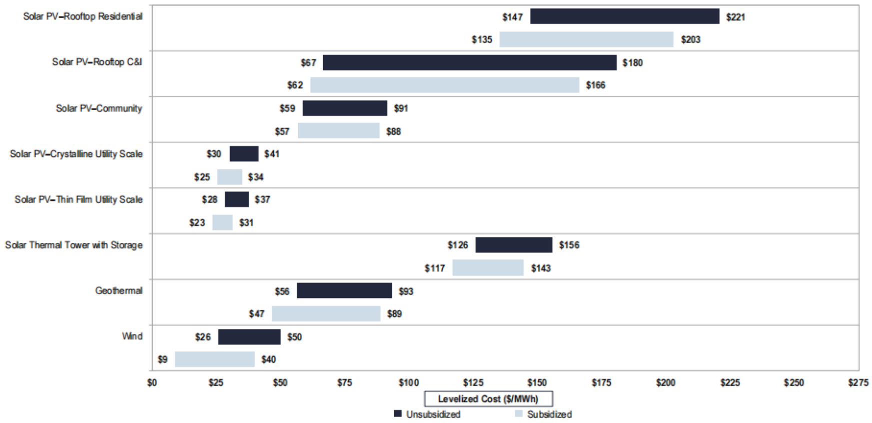


Figure 5.8: Subsidized vs Unsubsidized Renewable Energy LCOE Cost

Table 5.3 shows the cost of initial investment for a 1 Mega Watt data center, Table 5.4 gives per kWh cost in dollars of energy produced through coal vs energy produced through solar/wind. It shows that green energy has lower costs per kWh compared to brown energy resource. These costs are based on average costs in USA, and may vary based on countries and location. Table 5.5 gives the subsidy on green energy extracted from analysis given in figure 5.8 taken from **[38]**. These subsidies promote use of green energy while carbon tax imposed per ton of carbon dioxide produced due to production and consumption of brown energy source discourages its use.

Table 5.3. Renewable Energy Initial Investment Cost for 1MW Data Center

Component	**Cost ($)**
Wind Turbine (1MW)	$1,600,000
Solar Panels (1MW)	$1,300,000
Energy Storage (1MW, 2 hours)	$400,000
Backup Generators (1MW)	$200,000
Total Green Energy Investment	**$3,500,000**

Table 5.4. Coal (Brown) vs Solar/Wind (Green) Energy Cost per kWh

Energy Source	**Cost per kWh ($)**
Coal Energy	$0.10
Green Energy (Solar/Wind)	$0.04

Table 5.5. Subsidy on Green Energy vs Carbon Tax on Brown Energy

Incentive	**Amount**
Subsidies for Green Energy	30% ITC
Carbon Taxes (Coal)	$25 per ton CO2

After all these basic measurements and values, we can perform payback period analysis through comparing initial investment into deploying green energy resource for a 1MW data center and the energy savings it causes which ultimately saves on monthly billing and cuts overall expenditure in a data center. These cumulative savings, at certain point, become equal to the value of initial investment and after that the actual revenue starts to generate and data centers get benefited by the savings. This way green energy not only protects environment but also ensures economic growth of data centers. The graph in Figure 5.9 shows the payback period at point in time t.

Table 5.6. Payback Period Analysis for Investment in Green vs Brown Energy for 1MW Data Center

Category	**Coal Energy ($)**	**Green Energy ($)**
Initial Investment	0	3,500,000
Subsidies	0	-1,050,000
Net Investment	0	2,450,000
Annual Operational Cost	50,000	80,000
Annual Energy Cost	876,000	350,400
Total Annual Cost	926,000	430,400
Annual Savings	0	495,600
Payback Period (Years)	-	Approx 4.94

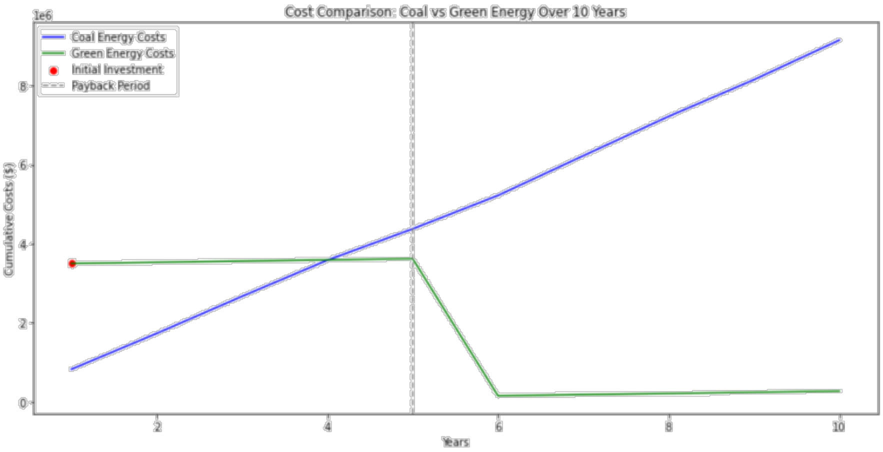


Figure 5.9: Payback Period for Renewable Energy

5.1.3 Savings Through Efficient Cooling Equipment

Investing in energy-efficient cooling equipment in data centers can lead to significant savings in both energy and maintenance costs, contributing to the overall economic sustainability of these facilities. Traditional cooling systems, like chillers, use a lot of energy because they rely heavily on mechanical processes to cool the air. This high energy consumption results in increased operational costs, which can be a major expense for data centers where keeping servers cool is crucial. On the other hand, newer, more advanced cooling technologies, such as air-side economizers and liquid cooling systems, offer more efficient ways to manage temperatures. Air-side economizers, for instance, use cooler outdoor air to reduce the need for mechanical cooling, cutting down on energy use. When the weather is cool enough, these systems can bypass traditional cooling methods altogether, saving even more on energy bills. Liquid cooling systems, which can cool components directly or through immersion, are especially effective for data centers with densely packed servers, as they can handle heat more efficiently than traditional air cooling. Natural cooling methods, like free cooling, also offer substantial benefits. By using the outside air or natural water sources to help dissipate heat, data centers can significantly cut down on the need for energy-intensive chillers, particularly in cooler climates. This not only reduces energy costs but also decreases maintenance needs, resulting in lower overall operating expenses. Embracing these energy-efficient cooling technologies helps data centers become more cost-effective and environmentally friendly, aligning with broader sustainability goals. In **[39]** authors discusses various cooling technologies and their economic metrics for data centers such as Vapor Compression Chiller, Free-Cooling, Indirect Free-Cooling and Evaporative Cooling, Trigeneration or combined cooling, heat and power (CCHP), Solar Cooling and Solar Photovoltaic. The study highlights that the most promising solutions for economic feasibility are the trigeneration and indirect free-cooling systems, both offering high internal rates of return and reasonable payback periods approximately 4 years with significant savings on cooling energy consumption. For analysis The Green Grid in **[40]** made a number of energy efficiency upgrades within a production data center and we will perform cost benefit analysis of all those upgrades. Upgrades made in existing infrastructure were Variable Speed Drives (VSDs) Installation, Upgrade of Older CRAH Units, Rack Airflow Management, Repositioning CRAH Sensors, and Temperature Setpoints Adjustment. All Computer Room Air Handlers (CRAHs) were retrofitted with OEM VSD fans capable of operating from 60% to 100% of capacit, older CRAH units were upgraded to newer, more efficient models with VSDs to improve airflow and cooling capacity, air dam baffle kits and blanking panels were installed in approximately 765 cabinets to improve isolation of hot and cold air aisles, temperature/humidity sensors were moved from CRAH inlets to the front of IT equipment racks. All these measurements are given in Table 14 for better understanding. The analysis in **[40]** shows that 963,000 kWh of energy was saved per year from VSD installations alone and 192,000 kWh from upgrading CRAH units which saved approximately $130,000 in one year. These upgrades had payback periods of minimum 1.3 years to 2.4 years with $4010,800 in ROI.

Table 5.7. Cooling Efficiency Upgrades vs Annual Energy and Cost Savings

Upgrade	**Unit Cost ($)**	**Quantity**	**Total Cost ($)**	**Annual Energy Savings (kWh)**	**Annual Cost Savings (USD)**
VSD Installation	3,000	24	72,000	1,261,440	151,372.80
CRAH Unit Upgrade	25,000	14	350,000	367,920	44,150.40
Rack Airflow Management	50	765	38,250	Included in above savings	Included in above savings
Sensor Repositioning	1,000	1	1,000	Enhanced efficiency	Included in above savings
Temperature Setpoint Adjustment	Minimal	-	Included in Labor cost	Enhanced efficiency	Included in above savings
Total	-	-	**461,250**	**1,629,360**	**195,523.20**

Table 5.8 given below shows the payback period analysis done through the initial costs, energy savings, cost savings on energy billing taken from table 5.6. It shows that all these cooling efficiency upgrades save massively on energy and annual cost reduction in billing and has payback period of as short as 2 years and 3 months. The replacement life of these components is more than 5 years, so data centers benefit from overall savings through these practices for the rest of the years.

Table 5.8. Cooling Efficiency Upgrades vs Annual Energy and Cost Savings

Improvement	**Initial Investment ($)**	**Annual Energy Savings (kWh)**	**Annual Cost Savings (USD)**	**Payback Period (Years)**
Variable Speed Drives (VSDs)	72,000	1,261,440	151,372.80	0.48
CRAH Unit Upgrade	350,000	367,920	44,150.40	7.93
Rack Airflow Management	38,250	Included in above savings	Included in above savings	-
Sensor Repositioning	1,000	Enhanced efficiency	Included in above savings	-
Temperature Setpoint Adjustment	Minimal	Enhanced efficiency	Included in above savings	-
Total	**461,250**	**1,629,360**	**195,523.20**	**2.36**

Similarly in another case study discussed in **[41]** at Google's data centers, the facility's cooling overhead was brought down from 1.4 to 0.5 of IT computer energy just by adding temperature monitoring, optimizing air vents, increasing temperature and humidity settings in CRACs, blocking curtains in cold aisles, and adding a new CRAC controller. All these improvements brought PUE from average 2.4 down to 1.5 and contributed to approximately USD 5500 of energy savings with a payback period of just 9.7 months at maximum. Google invested $25,000 to enhance the airflow in a room and minimize air conditioner usage. This expenditure on plastic curtains, air return extensions, and a new air conditioner controller led to annual savings of $67,000. Notably, the retrofit was completed without causing any operational downtime. In another case study by Schneider electric **[42]** capital cost analysis of air vs liquid cooled data centers was done which is given in table 15 and it was found that air cooling is typically less expensive to install and maintain but is less efficient and required more space. However liquid cooling has high initial costs but it is more efficient at removing heat especially in high density data centers. With the expenditures done and ROI analysis the payback period through all the savings was approximately 5.6 months.

Table 5.9. Air vs Liquid Cooled Data Centers Energy and Cost Savings

Aspect	**Air-Cooled Data Center**	**Liquid-Cooled Data Center**
Initial Investment	\$20,000,000 (for a 2MW data center)	\$19,540,000 (for a 2MW data center)
Annual Energy Consumption	17,520,000 kWh/year	-
Energy Savings	-	48% energy savings compared to air-cooled systems
Annual Energy Savings	-	8,409,600 kWh/year
Annual Cost Savings	-	\$1,009,152/year

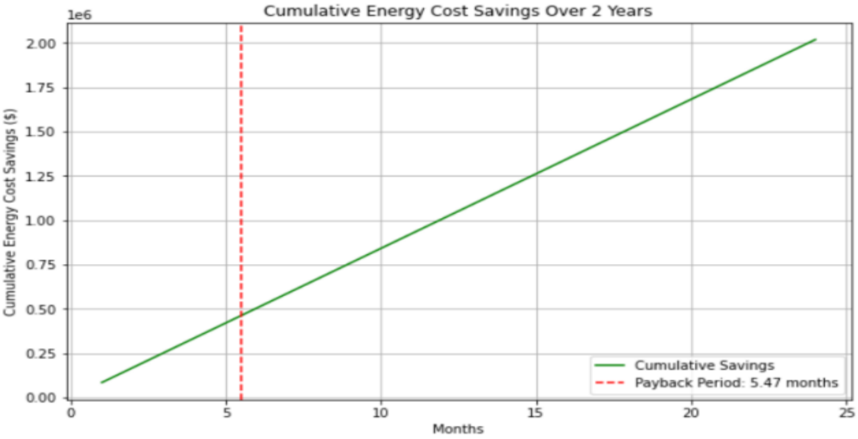


Figure 5.10: Payback Period and Cumulative Savings through Liquid Cooling

5.2 Energy Savings through Efficient Task Scheduling Techniques in Cloud: Discussion

Task scheduling in data centers is the process of assigning jobs or tasks to computing resources like servers, CPUs, or virtual machines. It ensures that workloads are distributed efficiently to optimize performance and resource utilization. It prioritizes tasks based on factors such as urgency, resource needs, and service level agreements (SLAs). The scheduler must also consider energy efficiency, aiming to minimize power consumption while maintaining performance standards. Techniques like Dynamic Voltage and Frequency Scaling (DVFS) adjust the power settings of processors based on workload demand, reducing energy usage during low-demand periods. Energy-aware task scheduling algorithms further enhance efficiency by considering the energy profiles of servers and directing tasks to those that can execute them with the least energy consumption. These algorithms incorporate predictive models to forecast workload demands and adjust schedules proactively. By optimizing resource allocation and reducing the need for over-provisioning, task scheduling helps data centers operate more sustainably.

5.2.1 Efficient Task Scheduling Techniques

5.2.1.1 Dynamic Voltage and Frequency Scaling (DVFS)

By modifying a processor's voltage and frequency in response to workload demands, dynamic voltage and frequency scaling, or DVFS, lowers power consumption. Because power consumption is related to the square of the voltage, reducing the frequency and voltage during periods of low utilization greatly reduces power consumption. Data centers frequently use this technique to balance performance and energy efficiency, particularly after hours. To effectively estimate workload demands and make real-time modifications that maintain acceptable performance while optimizing energy consumption, DVFS requires complex algorithms. When combined with workload prediction models, DVFS can dramatically reduce cloud service energy footprints.

5.2.1.2 Consolidation and Virtual Machine (VM) Migration

By combining several workloads onto fewer servers through server consolidation and virtual machine migration, idle servers can shut down or go into low-power modes. In order to achieve higher resource utilization and lower overall energy consumption, this strategy uses virtualization to transfer virtual machines (VMs) dynamically based on current resource utilization. By ensuring effective task distribution, virtual machine migration reduces energy consumption. Regular migrations, however, might result in expense and performance degradation, so it's important to weigh the advantages of consolidation against any potential drawbacks. By adapting to changing workload patterns, this technique minimizes energy consumption and maximizes resource utilization by handling demand spikes and maintenance without any downtime. In cloud scheduling, the ISN (Idle Sleep Notification) policy controls the transitions between various operational states according to work arrivals and idle times with the goal of optimizing server energy consumption. When a server is actively processing jobs under this policy, it starts off in the Busy mode. The server enters idle mode if there are no more jobs arriving, during which it waits for any new jobs to arrive. In the event that a new work comes in during this period, the server switches back to Busy mode to handle it. The server switches to Sleep mode in order to conserve energy if no jobs come in during the idle time. When the server wakes up from Sleep mode, it determines if there are more jobs in queue than a predetermined amount (N). When the cutoff is reached, then entering Busy mode, the server manages the burden. Otherwise, it stays in a low-power condition and keeps an eye out for new jobs. This policy minimizes idle time and takes use of sleep phases during periods of low demand to efficiently reduce needless energy consumption.

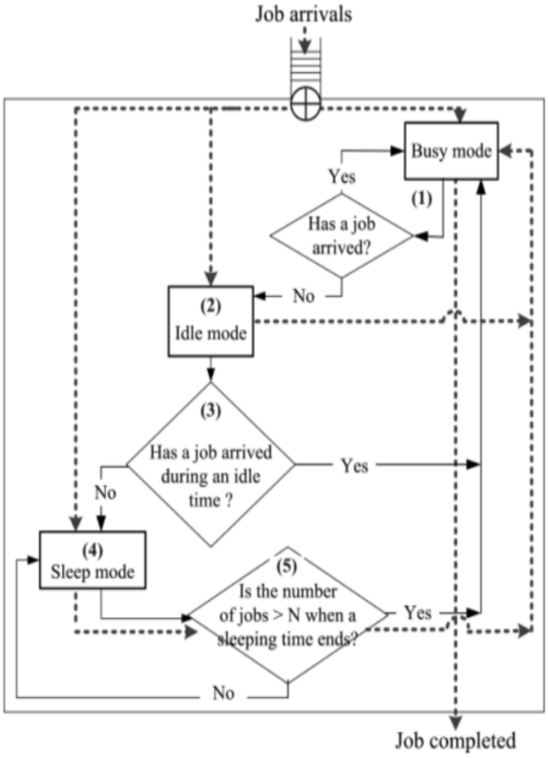


Figure 5.11: ISN Policy Framework

5.2.1.3 Energy-Aware Task Scheduling Algorithms

Task scheduling methods that include energy efficiency provide equal weight to performance and delay. In order to lower overall power consumption, these algorithms distribute workloads among servers according to energy profiles, workload characteristics, and thermal states. Min-Min, Max-Min, and energy-efficient genetic algorithms are a few examples. These algorithms drastically reduce power usage by scheduling jobs on servers that perform them with the least amount of energy consumption. They use machine learning or heuristics to optimize the allocation process, and they use predictive models to foresee workload demands and make proactive scheduling adjustments. Maximizing resource usage while reducing energy consumption and preserving service quality is the major objective.

5.2.1.4 Energy-Aware Load Balancing

In order to reduce energy usage, workloads are distributed among servers using energy-aware load balancing. In contrast to conventional load balancing methods that prioritize performance measures, energy-aware approaches take into account the energy profiles and current usage levels of servers. Through the assignment of workloads to servers that are operating at maximum energy efficiency, these methods lower total power consumption. Energy-aware load balancing decisions are informed by real-time monitoring and predictive analytics, which strike a balance between performance needs and energy savings. Preventing bottlenecks and making sure energy-efficient load balancing maintains data center dependability and efficiency while maximizing energy consumption and performance are challenges.

5.2.1.5 Green Scheduling

Green scheduling synchronizes work completion with times when renewable energy sources are most accessible, like during periods of maximum solar or wind power generation. By using renewable energy to power data center operations, this method lessens carbon emissions and dependency on fossil fuels. Green scheduling matches workloads with the availability of green energy, optimizing the energy mix and enhancing sustainability. It is necessary to integrate renewable energy generation with energy management systems and to monitor it in real time. One of the challenges is to prioritize the use of renewable energy while preserving performance indicators, and to make sure that green scheduling facilitates effective and sustainable data center operations.

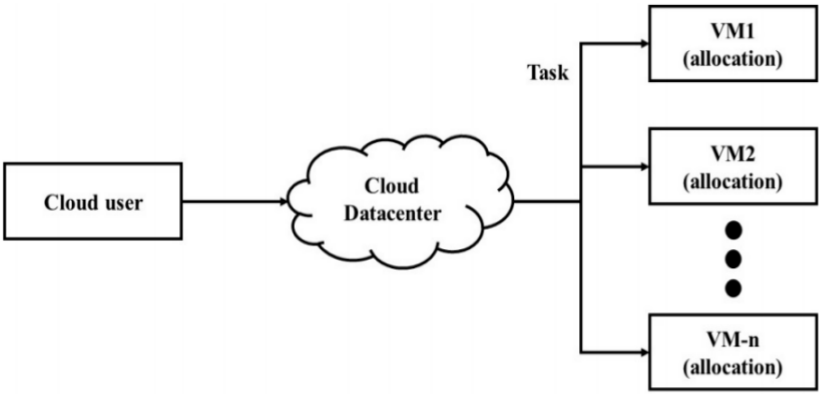


Figure 5.12: Virtual machine (VM) allocation in cloud data centers.

The primary goal is to maximize data center revenues while maintaining quality of service (QoS) and reducing carbon footprints. Server consolidation involves dynamically turning servers on or off based on incoming workload to enhance energy efficiency without compromising service quality. The article in **[45]** explores the significant power consumption issues faced by cloud-based data centers (DC) and proposes server consolidation as a viable solution. Server consolidation aims to increase average utilization by reducing the number of active servers, thus saving power. Their model considers factors such as power consumption during server startup/shutdown, idle and peak power usage, and renewable energy integration. Several experiments compare revenue generation and energy consumption with and without server consolidation. Results indicate that server consolidation significantly reduces power consumption and increases revenue, particularly in scenarios with varying server counts and traffic loads. Consolidation returns positive revenue by reducing idle power consumption and optimizing resource usage.

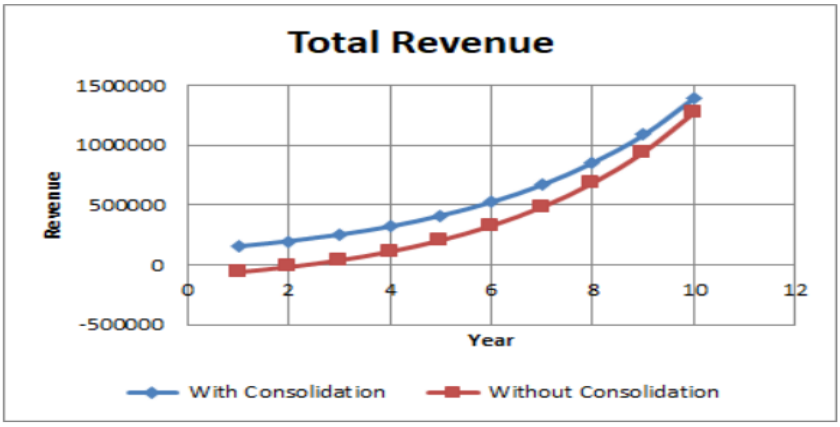


Figure 5.13: Revenue With vs Without Server Consolidation through Energy Savings

Chapter 6

Conclusion

It is important to raise understanding of energy efficiency from a young age in order to combat ignorance and promote a society that values conservation. People need to be made aware of the importance of energy conservation in all facets of life, even in small things like turning off the lights. Furthermore, the ingrained mentality that "buy more, pay less" wastes a lot of energy and needs to be challenged in order to encourage more environmentally friendly behaviors. The costs of an energy policy are frequently computed in standard policy analysis, but the benefits could be overstated or viewed too narrowly. Although a thorough cost analysis is essential, it does not give a whole picture of how a new policy will affect a state, tribe, or neighborhood. Clear decision-making is hampered when the advantages of energy efficiency and renewable energy policies are not fully represented or included in the final analysis. This can also keep policymakers in the energy, environmental, and business sectors from realizing and realizing the full potential benefits of these policies. This thesis is an extended work of previously published paper in ComTech 2023 “Data Centers Sustainability: Approaches to Green Data Centers”. After exploring various green strategies for enhancing data centers sustainability and performing cost-benefit analysis for them, we reach to the conclusion that green practices for data centers not only make them less harmful for environment but also benefit the owners massively through savings on energy and maintenance costs. The average payback period of all the studies discussed in this research didn’t exceed 5 years which shows that green infrastructure is a investment that produces positive ROI and generates revenue for the data centers contrary to belief that such investments produced negative cash flow. However, for small scale data centers investing in green infrastructure can be risky. In future work of this thesis we aim to find the impact of different energy efficient task scheduling algorithms on overall energy savings and costs of data center.

Bibliography

[1] International Energy Agency, "Data centres and data transmission networks," IEA. [Online]. Available: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks.

[2] U.S. Department of Energy, "Data centers and servers," Energy.gov. [Online]. Available: https://www.energy.gov/eere/buildings/data-centers-and-servers.

[3] Soergel, Bjoern, et al. "A sustainable development pathway for climate action within the UN 2030 Agenda." Nature Climate Change 11.8 (2021): 656-664.

[4] United Nations, "Transforming our world: The 2030 agenda for sustainable development," United Nations. [Online]. Available: https://sdgs.un.org/2030agenda.

[5] Fatima, Eeman, and Sidrish Ehsan. "Data Centers Sustainability: Approaches to Green Data Centers."2023 International Conference on Communication Technologies (ComTech). IEEE, 2023.

[6] Agarwal, Anup, et al. "Redesigning data centers for renewable energy."Proceedings of the 20th ACM Workshop on Hot Topics in Networks. 2021.

[7] Laganà, Demetrio, et al. "Reducing the operational cost of cloud data centers through renewable energy."Algorithms11.10 (2018): 145.

[8] Gough, Corey, Ian Steiner, and Winston Saunders.Energy efficient servers: blueprints for data center optimization. Springer Nature, 2015.

[9] Fawaz, AL-Hazemi, et al. "Dynamic allocation of power delivery paths in consolidated data centers based on adaptive UPS switching."Computer Networks144 (2018): 254-270.

[10] Asadi, Shaymaa Kamil Mwayesh. "GREEN COMPUTING AND ITS ROLE IN REDUCING COST OF THE MODREN INDUSTRIAL PRODUCT."Academy of Accounting and Financial Studies Journal25.6 (2021): 1-13.

[11] Tripathi, Rakesh, S. Vignesh, and Venkatesh Tamarapalli. "Optimizing green energy, cost, and availability in distributed data centers."IEEE Communications Letters 21.3 (2016): 500-503.

[12] Tripathi, Rakesh, S. Vignesh, and Venkatesh Tamarapalli. "Optimizing green energy, cost, and availability in distributed data centers."IEEE Communications Letters 21.3 (2016): 500-503.

[13] Uchechukwu, Awada, Keqiu Li, and Yanming Shen. "Energy consumption in cloud computing data centers."International Journal of Cloud Computing and Services Science (IJ-CLOSER)3.3 (2014): 31-48.

[14] Song, Zz, Xiaojing Zhang, and Clas Eriksson. "Data center energy and cost saving evaluation."Energy Procedia75 (2015): 1255-1260.

[15] Lin, Weiwei, et al. "A cloud server energy consumption measurement system for heterogeneous cloud environments."Information Sciences468 (2018): 47-62.

[16] Rodriguez-Martinez, Manuel, et al. "Estimating power/energy consumption in database servers."Procedia Computer Science 6 (2011): 112-117.

[17] Coles, H. C., Y. Qin, and P. N. Price. "Comparing server energy use and efficiency using small sample sizes. Lawrence Berkeley National Laboratory." (2014).

[18] Alkrush, Ahmed A., et al. "Data Centers Cooling: A Critical Review of Techniques, Challenges, and Energy Saving Solutions."International Journal of Refrigeration(2024).

[19] Zhang, Hainan, et al. "Free cooling of data centers: A review."Renewable and sustainable energy reviews35 (2014): 171-182.

[20] Ham, Sang-Woo, et al. "Energy saving potential of various air-side economizers in a modular data center."Applied Energy138 (2015): 258-275.

[21] Patel, Chandrakant D., et al. "Smart cooling of data centers."International Electronic Packaging Technical Conference and Exhibition. Vol. 36908. 2003.

[22] Zoie, Rădulescu Constanţa, Rădulescu Delia Mihaela, and Sipică Alexandru. "An analysis of the power usage effectiveness metric in data centers."2017 5th International symposium on electrical and electronics engineering (ISEEE). IEEE, 2017.

[23] Anderson, Sean F. "Improving data center efficiency."Energy Engineering 107.5 (2010): 42-63.

[24] Azevedo, Dan, et al. "Carbon usage effectiveness (CUE): A green grid data center sustainability metric." The green grid 32 (2010).

[25] Xu, Zhi-Wei. "Cloud-sea computing systems: Towards thousand-fold improvement in performance per watt for the coming zettabyte era." Journal of Computer Science and Technology 29.2 (2014): 177-181.

[26] Greenberg, Albert, et al. "The cost of a cloud: research problems in data center networks."ACM SIGCOMM computer communication review 39.1 (2008): 68-73.

[27] Koomey, Jonathan, et al. "A simple model for determining true total cost of ownership for data centers."Uptime Institute White Paper, Version 2.2007 (2007): 279.

[28] Uddin, Mueen, et al. "Classification of data center to maximize energy utilization and save total cost of ownership."International Review on Computers and Software7.5 (2012): 2106-2115.

[29] Rahmani, Rasoul, Irene Moser, and Mohammadmehdi Seyedmahmoudian. "A complete model for modular simulation of data centre power load."arXiv preprint arXiv:1804.00703(2018).

[30] An, Haiyun, and Xiaoyan Ma. "Dynamic coupling real-time energy consumption modeling for data centers."Energy Reports 8 (2022): 1184-1192.

[31] Azimi, Reza, Xin Zhan, and Sherief Reda. "Thermal-aware layout planning for heterogeneous datacenters."Proceedings of the 2014 international symposium on Low power electronics and design. 2014.

[32] Salom, Jaume, Laura Sisó, and María Sansigre. "Cost-effective renewable cooling on green data centres."Barcelona, Dec(2014).

[33] Bliedy, Doaa, Sherif Mazen, and Ehab Ezzat. "Cost Model for Establishing a Data Center."International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol 8 (2018).

[34] STAR, ENERGY. "3. Investment Analysis."

[35] Gough, Corey, Ian Steiner, and Winston Saunders.Energy efficient servers: blueprints for data center optimization. Springer Nature, 2015.

[36] “Astroids@home,” https://asteroidsathome.net/boinc/cpu list.php.

[37] Driving, Increasing Data Center Density While. "Down Power and Cooling Costs White Paper."Intel Corporation, June(2006).

[38] Doyle, Joseph, and Rabih Bashroush. "Case studies for achieving a return on investment with a hardware refresh in organizations with small data centers." IEEE Transactions on Sustainable Computing 6.4 (2020): 599-611.

[39] Star, E. "Energy savings from energy star-qualified servers." (2010)

[40] Zakarya, Muhammad. "Energy, performance and cost efficient datacenters: A survey."Renewable and Sustainable Energy Reviews 94 (2018): 363-385.

[41] International Renewable Energy Agency IRENA.Renewable power generation costs in 2020. eBook Partnership, 2022.

[42] Raphael Ebenhoch et al. “Comparative Levelized Cost of Energy Analysis”. 12th Deep Sea Offshore Wind RD Conference, EERA DeepWind’2015

[43] LAZARD. Levelized Cost of Energy Analysis. Version 15.0

[44] Salom, Jaume, Laura Sisó, and María Sansigre. "Cost-effective renewable cooling on green data centres."Barcelona, Dec (2014).

[45] Brey, Tom, et al. "Case study: the ROI of cooling system energy efficiency upgrades."White paper 39 The green grid: energystar. gov(2011).

[46] Google, “Google’s Green Data Centers: Network POP Case Study

[47] Bunger, Robert, Wendy Torell, and Victor Avelar. "Capital cost analysis is of immersive liquid-cooled vs. air-cooled large data centers."Schneider Electric White Paper 282 (2020).

[48] J. Hamilton, “Data Center Cooling Done Differently,” Perspectives, Oct. 2014;

[49] Wardat, Mohammad, et al. "Cloud data centers revenue maximization using server consolidation: Modeling and evaluation."IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2018.