University of South-Eastern Norway

www.usn.no

FMH606 Master's Thesis -2024

Electrical Power Engineering

Optimal Energy Management Strategies for Data Centers

Sandhya Bohara

Faculty of Technology, Natural sciences and Maritime Sciences

Campus Porsgrunn

University of South-Eastern Norway

www.usn.no

Course: FMH606 Master's Thesis, 2024

Title: Optimal Energy Management Strategies for Data Centers

Number of pages: 54

Keywords:

Student: Sandhya Bohara

Supervisor: Sambeet Mishra and Thomas Øyvang

Co-Supervisor: Chiara Bordin, The Arctic University of Norway (UiT)

External partner: Lede AS

Summary:

Data centers, the heart of the digital infrastructure, have been an integral part of daily life. It is expected that demand for cloud computing will rise due to the industry's rapid growth and AI scale-up. This thesis provides a comprehensive analysis of data center operations, focusing on scale and performance. It highlights how underutilized backup resources can be proposed as batteries to enhance the efficiency and sustainability of the power grid, potentially offering financial benefits to data center operators.

Key load balancing strategies such as time shifting, geography shifting, power load adjustment, and UPS energy storage are thoroughly studied, showing how they improve data center flexibility. The load curve analysis under different loading, as well as the corresponding load factor, reveals the energy resources that can be optimized. To maximize profit, an Excel model takes into account workload prioritization, load curves, revenue generation, and electricity costs. Grid optimization and location feasibility are guided by load flow analyses and simulations in DIgSILENT PowerFactory.

Future research could improve these optimization strategies, explore their application in diverse contexts, and incorporate AI and machine learning for greater efficiency. Contingency analysis is suggested for further studies on grid reliability, aiming to balance energy consumption and enhance grid stability.

The University of South-Eastern Norway takes no responsibility for the results and conclusions in this student report.

Preface

The thesis was done under the Electrical Power Engineering department as part of the FMH606 Master's Thesis course under the guidance of supervisor Sambeet Mishra, Thomas Øyvang, and Chiara Bordin at the University of South-Eastern Norway (USN) written in the final semester in the spring of 2024. I would like to express my heartfelt thanks for their valuable guidance and support during the semester. It was a highly valuable experience for me to work under their close supervision.

The initial concept of the thesis was developed through discussion with Lede AS and focuses on understanding, exploring, and developing innovative methodologies for optimizing data center operations and integrating them into power systems.

I would like to express my gratitude and love to my family for their unwavering support and affection.

Porsgrunn, 29th May 2024

Sandhya Bohara

Contents

Nomenclature

Acronyms

DCs	Data centers
TIA	Telecommunications Industry Association
BMS	Building Management System
CS	Cooling System
FSS	Floor Standing Cooling Systems
FCS	Free Cooling Systems
PUE	Power Usage Effectiveness
TIA	Telecommunications Industry Association
AWS	Amazon Web Services
UPS	Uninterrupted Power Supply
RES	Renewable Energy Resources
DR	Demand Response
VPP	Virtual Power Plant
CRAC	Computer Air Room Conditioning
PV	Photovoltaic
ISO	Independent System Operator
FCR-N	Frequency containment reserves for normal operation
FCR-D	Frequency containment reserves for disturbances
PDU	Power Distribution Units
SLA	Service Level Agreement

List of Figures

List of Tables

1 Introduction

This chapter contains the background of this thesis, objectives and scope, methods used, and an overview of the thesis.

1.1 Background

Data centers (DCs), the heart of the digital infrastructure, have been an integral part of daily life. In 2018, the Norwegian government introduced the world's first national strategy for establishing data centers [1]. According to the data in the report [2], as of December 2023, the maximum installed capacity of the data center industry in Norway is 501 MW, of which 150 MW is currently in utilization. In 2023, the data center used one percent of the electricity produced in Norway and is expected to rise to 1.9 percent by 2028 due to the industry's rapid growth and AI scale-up.

Data centers store, process, and transfer large amounts of data thus managing and maintaining the data center infrastructure is an important part of its operation. They are the big energy consumers. For instance, a hyper-scale data center can consume as much energy as 80000 households which causes more pressure to make them more sustainable. The data center's large energy consumption increased operating costs and had an adverse effect on the environment. The worldwide electricity consumed by DCs has increased by 56% from 2005 to 2010, accounting for 1% of the global electricity consumption in 2020, and is predicted to be compromising 13% of the energy consumed in the world by 2030 as the world becomes more digitally connected. Datacenter loads can be divided into IT and non-IT loads. The IT load Includes Servers, Storage systems, networking equipment, firewalls, and security appliances. The non-IT load includes HVAC, Power distribution systems, UPS, Generators, Lighting, Facility security systems (access control, surveillance, cameras, and alarms), and BMS (Building Management System). In DCs, nearly 52% of the electricity is consumed by the IT equipment, 48 % by the non-IT which includes 38% by the cooling system (CS), and 10% by the remaining equipment.[3]

The cooling system accounts account 38 % of a data center’s energy consumption since thousands of servers need to be cooled to work efficiently. How well a data center cools its servers determines the capacity of a data center. Over the past decades, cooling technology has improved significantly and most of the large data centers have replaced the old air-conditioning-like systems that cool the entire room with in-row or rotodynamic heater-based cooling designs. This design draws the heat emitted from the servers away by fans which is subsequently cooled with water or a refrigerant. The few other cooling technologies are floor-standing cooling systems (FSS), and free cooling systems (FCS) with their value of power usage effectiveness (PUE). The PUE is the ratio of the total energy used by a DC, to the energy delivered to IT equipment.

Developing the optimal energy management strategies that can benefit data centers is crucial. Load balancing is an essential part of the operation of data center management, the system that makes sure the workloads are distributed evenly and effectively among the available resources of the data center for optimal performance and reliability. Optimizing data centers' energy use could greatly impact how the power network is loaded, contributing towards sustainability, cost-effectiveness, and overall operational efficiency[4]. A practical load-balancing mechanism is required to balance the energy consumption on the grid as well [3].

Certain governments and regulators impose sustainability standards on newly built facilities. Thanks to this progress, investors now could support data centers in obtaining a carbon-free energy supply [5]. GreenWare, a middleware system introduced in the paper [6] optimizes the use of renewable energy in a network of distributed data centers despite intermittent renewable energy supplies, time-varying electricity prices, and shifting workloads while meeting Internet service operators' costs. It addresses two key questions for cloud-service operators: (1) how to dynamically distribute service requests across data centers based on local weather conditions to maximize the use of renewable energy, and(2) how to do so within their operational budgets. The Internet service provider can utilize this approach in their cloud-scale data centers which will contribute to mitigating the negative environmental impacts such as CO2 emissions and global warming because of rising energy consumption.

This paper [7] investigates about optimal power cost management from the use of UPS units as energy storage devices. Lyapunov optimization has been used to develop an online control algorithm that maximizes UPS utilization while reducing the time average electricity bill of the data center. This algorithm operates independently of workload and electricity cost statistics, making it appropriate for unpredictable workload and pricing scenarios.

Energy efficiency has restricted the growth of large enterprises, such as data centers, which run on many servers and consume a significant amount of power. Thus, data center-based energy efficiency will become an unavoidable trend. This paper [8] presents a model for maximizing the benefits of data centers and structuring them to reduce energy consumption. The model considers various factors, such as electricity price, renewable power generation, service rate, and Quality of Service (QoS), to balance benefits and costs. The clustering algorithm is used to optimize data center task scheduling, reducing energy consumption, and improving efficiency. It highlights that in large-scale data centers, data mining is used to optimize resource allocation to user tasks. Users can select cloud resources based on their needs, resulting in an equal balance of resource supply and consumption.

1.2 Novelty and Contribution

Several energy management strategies were implemented within the data center operations and infrastructure optimization. This research takes an original approach, combining simple load prioritization modeling techniques for profit maximization based on their priority, cost, and potential revenue impact while providing a better comprehension of data center operations within power systems.

This thesis explores the relationship between data center operational models and power grid models, which is currently underexplored in the literature.

1.3 Objective and Scope

The main goal of this thesis is to document the development of innovative methodologies for data center operations modeling and their integration into power systems and to present a detailed analysis of the value of data center flexibility through extensive analysis.

1.4 Approach and Methodology

To understand the complex nature of data center operations and potential load-balancing mechanisms, a thorough review of the literature and research has been done first. which involved examining and understanding current modeling approaches for both data centers and power networks.

An effective strategy will be developed where workloads are prioritized to effectively manage varying demands among computational resources. By giving priority to critical or high-demand workloads, data centers can ensure that important tasks are completed promptly and efficiently, even during periods of peak demand or resource shortage.

Excel (Microsoft Office 365) will be used to create a model, solve the optimization problem, and produce the data graphically. Following the creation of the model, the developed strategy is integrated into the industry-standard IEEE nine-bus power system, and in-depth analyses are conducted to evaluate the benefits of data center flexibility

1.5 Thesis Overview

This thesis consists of four main chapters. Chapter 1 gives an introduction about the data centers with their loads and the existing modeling approaches implemented for data centers' optimal operations, the relative background, and the objectives of the thesis. Chapter 2 contains the types of data centers, an overview of load management strategies, how data centers can be operated as the battery, a comparative study of data centers, and an analysis of the flexibility and potential of data centers. Chapter 3 will discuss the profit optimization problem and the simulation that will integrate the result from the problem. Chapter 4 covers the overall conclusion and gives the outlook for possible future research and expansion on the topic.

2 Datacenter as a Load

To effectively treat a data center as a load and optimize its energy efficiency, it is critical to understand the various types of data centers, energy management strategies, and the flexibility and potential of these facilities.

In this chapter, the in-depth background and understanding of these components will be presented.

2.1 Types of Data Centers

The Telecommunications Industry Association (TIA) has established the TIA-942 standard for data centers that covers critical aspects, including architecture, safety, and physical centers based on infrastructure security, redundancy, and environmental design. The four-level rating data have been shared as shown in Table 1.

Table 1: Tier-level of data centers [9]

Tier Level	Description	Features	Downtime tolerance
Tier 1	Basic site infrastructure	UPS, dedicated physical area for IT system, cooling equipment, backup power generator	No redundancy
Tier 2	Redundant component site infrastructure	Additional cooling components like chillers, exhaust pumps	Partial redundancy
Tier 3	Concurrently maintainable site Infrastructure	Higher data redundancy, allows equipment maintenance without system shutdown, low annual downtime	High redundancy(N+1) for cooling and power
Tier 4	Fault-tolerant infrastructure	A physically isolated system, prevents disruptions from planned and unplanned events, very low downtime	Complete redundancy(2N+1)

Depending on the size and management of the data centers there are several types of data centers, and they are explained below:[9, 10]

a) Managed Data centers
They are the company or business that provides third-party services directly to the clients, including computing, data storage, and other services. They lease the infrastructure and services to those who want to optimize the overhead by cutting the cost of managing the data center.

b) Colocation data centers

It is also known as the multi-tenant data center. This facility provides proper data center components including the power, security, cooling, and networking components to the clients willing to rent it.

The colocation data center is the best option when one wants to reduce maintenance costs and monthly fixed expenses. Many industries including banking, healthcare, government agencies, and manufacturing companies have been taking advantage of this type of data center.

The demand for this facility is very high and customers anticipate constant uptime, a large amount of bandwidth, and instantaneous access to data.

c) Hyperscale data centers

They are the largest data centers with extensive infrastructure. For instance, a typical hyperscale data center can house roughly 5000 servers in less than 10,000 square feet or anything above 200 MW. Large hyperscale data centers are operated by corporations such as Amazon, Google, and Microsoft, primarily for cloud computing and big data storage which owns more than half of all the hyperscale data centers.

d) Cloud-based data centers

They are known as distributed data centers run by third-party managed service providers such as Amazon Web Services (AWS) or Google Cloud. Through cloud services, these data centers enable customers to rent space and infrastructure as needed, enabling the quick provision of virtual resources. Hence, they are known for their flexibility and scalability.

Cloud-based data centers are cost-effective as they cut off the cost of the essential components of the data center: data center staff, support infrastructure, core hardware, and physical location.

e) Edge data centers

Edge or Micro data centers are the smaller data centers that process, analyze, and act on data in real time and are located close to end users. Through the processing of data services as close to end users as possible, this kind of data center allows organizations to minimize communication delays known as latency, and enhance the customer experience.

f) Enterprise data centers

This kind of data center is typical in the technological industry built on or off-site and is utilized internally by a single company. The key benefits are improved security and customization however they are expensive to build and have high maintenance costs.

Companies with this type of center, in case of any disaster, can isolate the business operation with the data operation.

2.2 Overview of Load Balancing Strategies

Compared to the power demands in commercial office spaces, which range from 50 W/m2 to 110 W/m2, the data center's total power demand density is found to be between 120 and 940 W/m2. This demonstrates clearly that data centers are highly energy-consuming structures with significant energy-saving potential, making them an ideal focus of energy-saving initiatives. [11]

Data center's management behavior is essential to achieve the long-term goals of the business because it directly impacts the operational sustainability which is the behavior and risks besides the infrastructure design. As a result, one of the most important aspects of tier classification is operational sustainability.[12]

Considering the different demand management potentials for IT and non-IT loads, the organization can increase operational efficiency, reduce costs, and peak load, and improve the overall user experience. Thus, it requires the integration of cutting-edge technologies, strategic planning, and continuous optimization and monitoring.

Some of the demand-side management strategies for the data center loads are:

- Load Shifting
- Energy Efficiency Measures: Equipment efficiency, virtualization, and Energy Efficient HVAC System
- Renewable Energy Integration or Green Data Centers or Net zero energy consumption DCs.
- Energy Storage system: Operating data center as Virtual Power Plant(VPP) , frequency regulation and UPS capacity utilized

Some of the widely implemented load balancing strategies to make data centers more flexible are Time shifting, Geography shifting, adjusting power load, and UPS energy storage systems.

- Firstly, the time shifting is to delay the workloads that do not require immediate treatment for example by shifting this workload to the time when the renewable generation is at its peak.
- Secondly, geography shifting is shifting the world load/task to the region where the renewable generation is highest.
- Thirdly, the cooling system temperature adjustments allow for flexible power load adjustments. The server's functionality won't be significantly impacted.
- Fourth, by optimizing the charge and discharge time of the UPS energy storage system which involves charging during the peak time of the energy generation and discharging during the period of peak load thus providing the reserve and operational flexibility for the power system. [13]

The quantity of data processed in the data center is directly proportional to the amount of electricity used. Transferring the process request to another data center gives greater flexibility in the amount of power needed in the data center. If another data center with the capacity to handle higher loads has lower electricity costs, this method can also be applied there. However, this geographical shifting is something that only large-scale data centers can currently do because even though the workload is small in terms of computation, they still need large data to be retrieved, and it is not possible to shift all workload in the case of small data centers, the type and location of the data center are the aspects that determine which solution the data center could eventually apply. As a result, different kinds of data centers will employ various strategies to raise or enhance energy efficiency and flexibility.

Another important piece of information is the data's location. Grid congestion may result, for example, from data centers situated in densely populated areas. However, if it's located in more rural areas, they will experience bandwidth or grid connection issues.

2.3 Data Center as Virtual Power Plant (VPP)

Although many data centers have their redundancy systems, data centers still rely on the electric grid for their power requirements. Therefore, the data center won't be impacted if there is a problem with the city's electrical grid. Since data centers need to be online, always accessible, and have almost no downtime. The average cost of an unplanned power outage in the data center is $8, 851 per minute [14] .

However, can the data center energy resources such as Uninterrupted Power Supply (UPS) systems, generators, and Renewable energy resources (RES) be used as backup systems if there is stress on the grid, such as during periods of peak demand or fluctuations in renewable energy? Most of the time, data center energy resources are underutilized so by utilizing them, data centers can take part in Demand Response (DR) as a virtual power plant (VPP) that can contribute to the overall efficiency and sustainability of the power system grid and potentially provide financial benefits to the data center operators. In this way, a data center can support grid stability and decarbonization while simultaneously enhancing its resilience and sustainability.[14, 15]

As explained by the operational approach of combining demand-side management and on-site energy resources in the paper [16] because of the underutilized energy resources and the high load flexibility, data centers are the ideal choice for operating as VPP and also showed that operating the data center as a VPP not only contributes to the data center loads but also serves to the grid by selling its underutilized energy resources.

In Figure 2.1, the data center VPP consists of server load and computer air room conditioning (CRAC) units as a data center load. The onsite energy resources include battery storage, photovoltaic (PV) array, and backup generators fueled by natural gas. The VPP energy management system coordinates its load and energy sources with the Independent System Operator (ISO) through communication to participate in the DR program.

The process of operation of the energy management system (EMS) of the data center starts with the collection of generation and load demand information which is explained below:

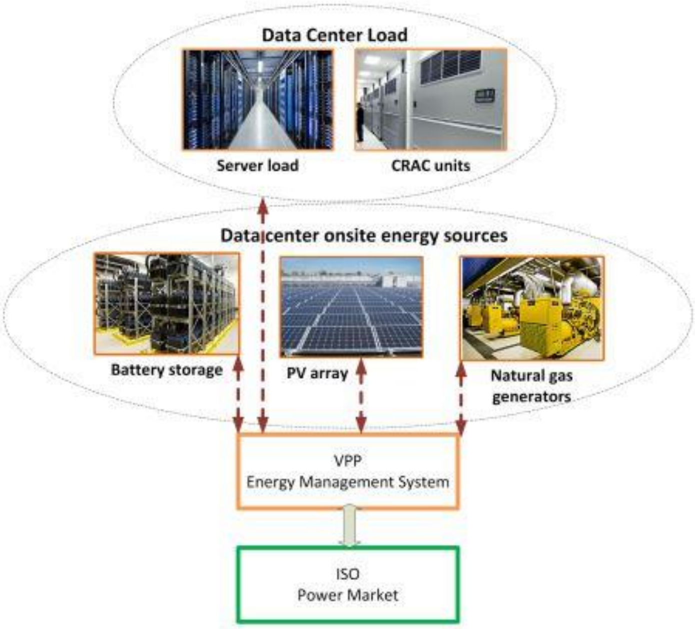


Figure 2.1: Data center operating as VPP [17]

- The data center energy management system (EMS) obtains a demand response (DR) signal which is a combination of generation and load demand from ISO and participates in a day-ahead DR program. Here, a real-time data monitoring module is used to collect the data from all the available energy resources in the data center.
- Once the data has been collected it is then sent to the forecasting module which uses the advance forecasting module to forecast next-day generation and load demand.
- The forecasted data are then used to calculate the net load which is the data center load minus renewable energy generation which is PV array generation.
- With the net load information, the amount of excess energy from resources or the deficit energy required for the data center load can be determined. [16]

In contrast with operating its generators, operating the data center as a Virtual Power Plant (VPP) and participating in demand response (DR) when network congestion conditions are taken into consideration saves a significant amount of money.

2.3.1 Primary services related to frequency regulation

Considering the redundancy requirement of the data center, its battery system is intentionally oversized. Hence, the data center can take part in primary services that are energy-non-intensive without any impact on its daily basis operation.

The paper [18] has suggested a novel approach of taking advantage of the bidirectional operations capabilities of the uninterruptible power supply (UPS) systems, thereby enabling them to provide dynamic power response from their battery systems. Even though the potential revenue from the primary regulation service is minimal compared to the electricity costs of the data center, it is nevertheless noteworthy because it has no impact on the daily basis operation of the data center.

Primary regulation services are found to be divided into two types namely a normal operation reserve and a disturbance reserve. Normal operation reserves are designed to handle the small variation in frequency and must be activated in minutes. Disturbance reserves are intended to handle sudden fluctuations in the ratio of generation and consumption and the activation of it must be in seconds. In a Nordic power system, the primary frequency regulation is categorized as the Frequency containment reserves for normal operation (FCR-N) and Frequency containment reserves for disturbances (FCR-D). The grid frequency range for FCR-N, a bi-directional normal operations reserve, is between 49.90 and 50.10 Hz. On the other hand, the only upward-regulating disturbance reserve, or FCR-D, begins to operate at 49.90 Hz and should reach full activation at 49.50 Hz. It’s been mentioned, that even though the frequency deviates from the nominal limit, the large deviation is rare which means FCR-D is activated but does not use its power completely.

The data center's redundancy level and implemented topology have a significant impact on exactly how much excess capacity is available. Data centers typically use the topologies N, N+1, 2N, and 2(N+1) where N stands for the basic distribution and capacity pathways. A redundant power delivery path is necessary for a data center to be classified as tier 3 (or higher) by the Uptime Institute, which means that at least a 2(N+1) UPS topology should be implemented[18].

Figure 2.2 shows four different topologies where 2(N+1) topology has two independent and redundant power delivery paths supplying critical loads of 3 MW.

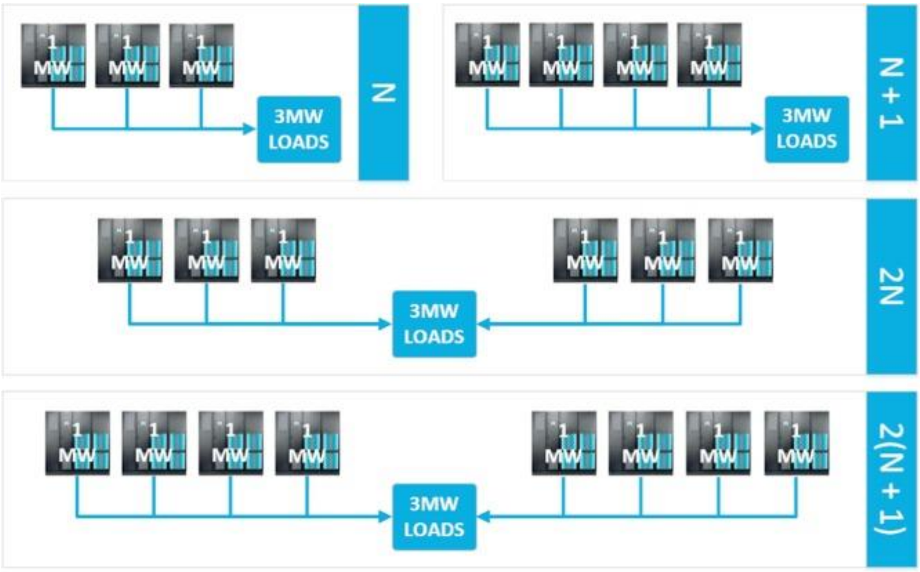


Figure 2.2: Different types of UPS topology in data centers[18]

Considering the 10 min of autonomy time (amount of time a UPS can feed power to the loads without any input power from main regulation supply), in 2(N+1) with 8 numbers of 1 MW

UPSs, total amount of energy in the battery systems is 1333 kWh feeding the 500 kWh (3MW*10min) load with excess energy of 833 kWh (5MW*10 min) in the battery system. The resultant excess energy thus can be utilized for the primary regulation during datacenter islanding and Dynamic upward regulation.

In Figure 2.3, the frequency-controlled breaker is used after the grid connection point. When the frequency drops below the predetermined threshold, the breaker will open, and the UPS will feed the critical loads until the generator (GENSET) is ready to feed and the data center will become a step-activated reserve and be eligible to take part in upward regulation.

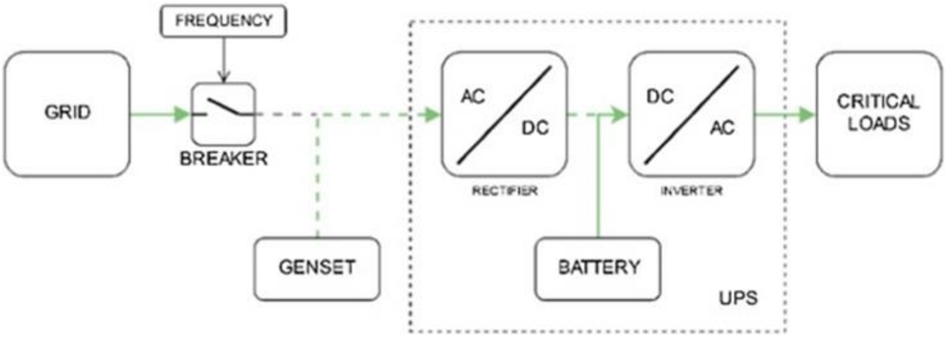


Figure 2.3: Data center Islanding [18]

In Figure 2.4, with the help of the battery system, the UPS system controls the power consumption from the grid. In case of frequency disturbance, the UPS system discharges energy from the battery, and depending upon the regulation need and power consumption of the load, it will either be consumed entirely by the on-site load or fed back to the grid if the regulation need is higher than the power consumption by the loads. Here, the regulation power is not limited to load power, and the UPS capacity is fully utilized.

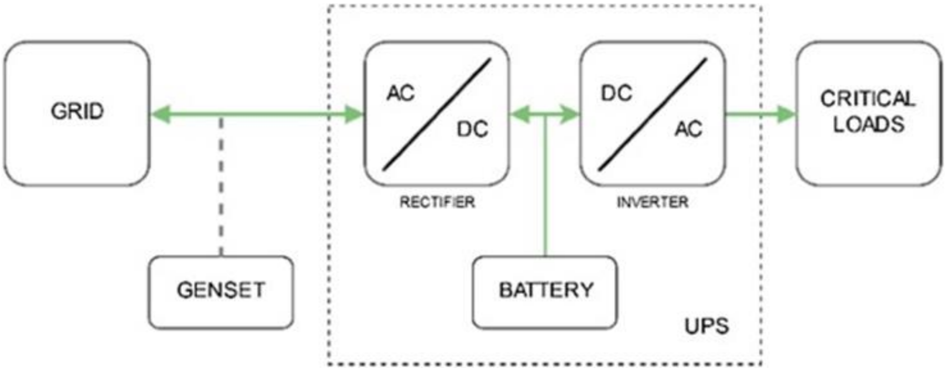


Figure 2.4: Dynamic upward Regulation

2.4 Comparative Study of Different Data Centers

This chapter covers the comparative study of data centers based on tier levels, focusing on redundancy, scale, and performance.

2.4.1 Comparison between different tiers of data center

The data center Tier classification levels were developed by Uptime Institute more than 25 years ago, and they continue to be the international standard for data center performance today [12]. It rates the data center and provides the certification into four levels based on the following few factors:

- Service availability and uptime guarantees.
- Redundancy levels which define the process of duplicating critical components and keeping them as backups and fail-safe in case of planned or unplanned disruptions.
- The state of cooling and power infrastructure.
- Staff expertise and maintenance protocols particularly the ability to handle concurrent maintainability.
- Service cost.
- Operational sustainability and the data center's ability to meet long-term business goals.
- The time a facility requires to set up a new client.
- Data center security levels.
- Carrier neutrality. [19]

From tier 1 to tier 4 in the Figure 2.5, the tier is progressive. This progression doesn't mean that tier 4 is better than tier 1; rather, it just shows that different businesses operate at different tiers.

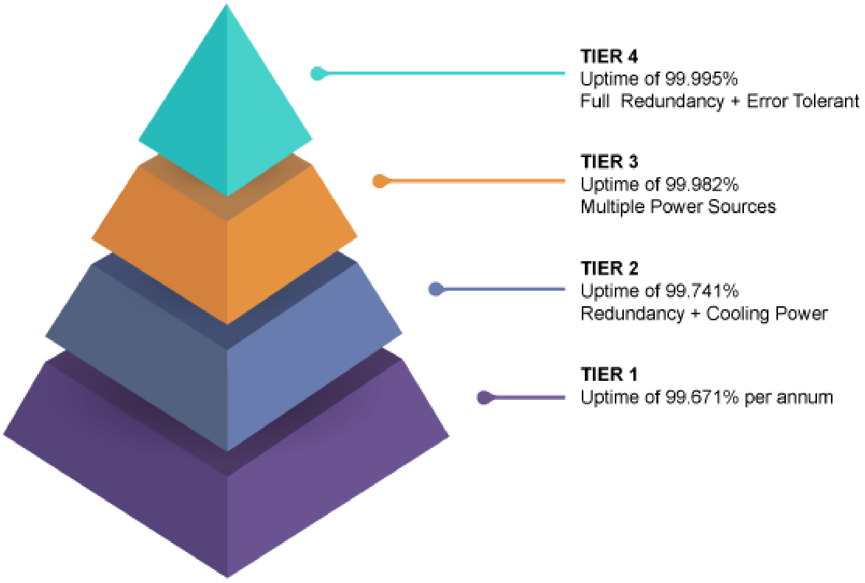


Figure 2.5: Data center classified tier level [20]

Typically, the two primary considerations when choosing a tier are cost and uptime. Paying for a level 3 data center when a cheaper facility would do the job is a waste of money. Likewise, setting up at a tier 2 facility when you require higher uptime can impact your revenue, productivity, customer satisfaction, and reputation. Most clients favor tier 3 data centers due to their long-term financial stability and operational sustainability, making them the most widely chosen standard across various industries.

Tier 4 with an uptime of 99.995% and 26.3 minutes of annual downtime has several independent and physically isolated systems acting as components of the redundancy and distribution path. It implies that the IT operation will continue uninterrupted even in the event of a piece of equipment failure or a disruption in the distribution path. Hence, it requires continuous cooling to make the working environment stable.

One thing to note is Tier 3 data centers offer a high level of redundancy but are not designed to be fault-tolerant compared to Tier 4 which are designed to tolerate the fault. One of the many reasons Tier 4 is superior to Tier 3 is the fault-tolerant infrastructure, multiple paths available for maintenance and replacement of equipment or power and cooling systems, so there's no need for shutdowns although the cost to build and operate it will be high [21].

Most clients have their checklists with long-term business goals and don't choose data centers based solely on tier classification. They will choose the one only if the checklist is satisfied. If the checklist focuses majorly on power, fault capabilities, maintenance, and redundancy, then the data center service provider must complete the process of obtaining the higher tier certificate for it to appear trustworthy. The most crucial is how the data center is kept up and maintained, as well as what kind of controls and checks are executed which are determined by the service provider's experience rather than by certification. Between tiers 3 and 4 of data centers, the choice will depend on the specific needs and budget of the business or clients.

2.4.2 Comparison based on Scale and Performance

The comparison of data centers is done based on scale and performance using the three factors of capacity and power supply, UPS and redundancy, and energy efficiency. [22]

Capacity and power supply: This measure shows how much IT infrastructure a data center can hold at its maximum capacity. It is represented in KW or MW depending upon the capacity of the data centers**.**

The overdesign of the power supply system and the overestimation of the power demand from utility results from oversizing HVAC and UPS equipment compounded by safety factors, and other factors. So, before designing the power supply for the HVAC and IT equipment, compare the peak loads in the data centers.

UPS and Redundancy:

This is one of the most important indicators for measuring the performance of any data center which measures the level at which the IT equipment can operate without any interruption.

The data center directly depends on the power utility as the main source of power supply, UPS as the power conditioner, and batteries and generators as backup power supply sources. The batteries will serve as an important source of power during a power outage for about 15 to 30 minutes, providing enough time to preserve data and properly shut down the IT equipment. The diesel generator will then start up and run until its main power source comes back online. A low load factor can lead to inefficient UPS energy consumption. For instance, data center 2 measured UPS load factor ranged from 11% to 15% and the efficiency was only 53-63%. In contrast, Data Center 1 showed better performance, with a load factor of 41% and an efficiency of 85% [11].

To fully meet the requirements of contingency planning, data center owners typically install N + 1 or 2N redundant power systems. These systems consist of a composite dual utility feeder from separate substations connected to on-site power conditioning equipment and UPS. So, if the data center has not employed any redundant UPS, it demonstrates poor redundant capability.

In some major scenarios such as unexpected failures, routine testing and inspections, and planned maintenance for the replacement of aging equipment or components, the several systems that help to optimally operate the IT equipment are disrupted. Thus, a data center that can support continuous uninterrupted operation of IT equipment is considered to have high availability. This availability is then represented by the Tier level of that data center which has been discussed in the previous chapter and the uptime; a data center's guaranteed annual availability which is expressed in percentage. A service level agreement (SLA) is a contract that guarantees uptime**.**

Energy Efficiency:

Power Usage Effectiveness (PUE), which evaluates how well a data center uses electricity, frequently serves as an indicator of efficiency. PUE is calculated by dividing the total load; the total electrical capacity the center can use by the IT load; and the electrical capacity available for IT equipment. It simply means it shows how much of the center's total electrical capacity is used for IT equipment. Lower values indicate higher efficiency. It typically lies between 1 and 2. A value closer to 1 indicates a center is designed and operated for higher efficiency.

$$\text{PUE} = \frac{\text{Total Load (total electrical capacity data center can use)}[\frac{\text{Kwh}}{\text{year}}]}{\text{IT load (the electrical capacity IT eqipment can use)[KWh/year]}}$$

The major processing load in data centers is IT equipment. A larger contribution from IT equipment to the total energy use indicates a better overall energy performance of the infrastructure and the supporting systems and a lower percentage of energy consumption from HVAC equipment indicates more efficient design and operation**.**[11]

The other energy efficiency measurement metric is the ratio of HVAC energy consumption to IT equipment energy consumption which is called metric 2. It ranges from 0.28 to 1.51 and on average is found to be 0.77.

$$\text{Metric 2} = \frac{\text{HVAC energy consumption }[\frac{\text{Kwh}}{\text{year}}]}{\text{IT equipment energy consumption [KWh/year]}}$$

For example, if the HVAC system required 0.84 kWh of electricity to remove the 1KWh heat that the IT system produced, the metric has the value of 0.84 indicating the effective use of energy in that specific data center. However, if the value falls out of the range, there could be several reasons behind it such as large space, placing of the thermostat, and room temperature variation but the major reason is the oversizing of the HVAC equipment.

2.5 Flexibility Potential of Data Center

Data centers have the option of using grid power or power produced on-site. The flexibility of the data center is increased when on-site power is utilized as a backup or primary power source because it gives more control over resource utilization. Nonetheless, having two independent power sources generally is a good idea to increase the data center's flexibility and stability. Efficiency and control over electricity, water, and carbon emissions are necessary for the flexibility to shift data center operations toward more sustainable practices. Improvements in sustainability are facilitated by the flexibility provided by smart liquid cooling, efficient infrastructure, and renewable energy sources. All the demand-side management (DSM) strategies also require load flexibility and data transfer capability.

2.5.1 Load profile with a different loading percentage

The daily load profile of the forecasted netload of the 100 MW Tier 4 data center based in Texas is [16] considered as a base for creating the load profile with different loading percentages. The calculation for total energy consumption, average load, and load factor for different percentages of loading can be seen in Appendix A. The data center is shown in Figure 2.6 , with a 50 MW solar plant, 50 MW wind power plant, and 192 MWh flooded lead acid battery providing backup for an hour long.

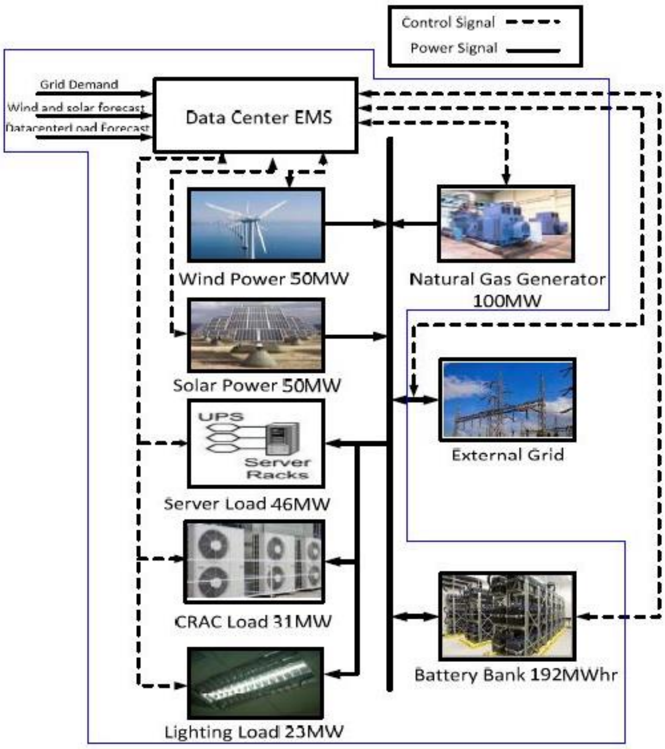


Figure 2.6: 100 MW data center with its generation resources and load AC coupled [16]

The daily load profiles of the data center for different loading percentages are visualized in Figure 2.7. In the base-load profile which is considered as 0 % loaded, the data center has a load factor of 53 % while at full loading, it was found to be 81 % loaded as shown in Table 2 and it indicates that data center resources such as servers, storage, and network bandwidth have all been utilized, resulting in an effective distribution of resources but still having some capacity available to handle the workload or data.

Load factor, also known as capacity factor, is a measure of how much of the maximum capacity of a system or facility is being utilized over a specific period. In the context of a data center, the load factor indicates how much of the data center's maximum power capacity is being utilized by the servers and other equipment at any given time. The load factor is calculated using the equation below:

$$\text{Load factor} = \frac{\text{Average load}}{\text{Peak load}}$$

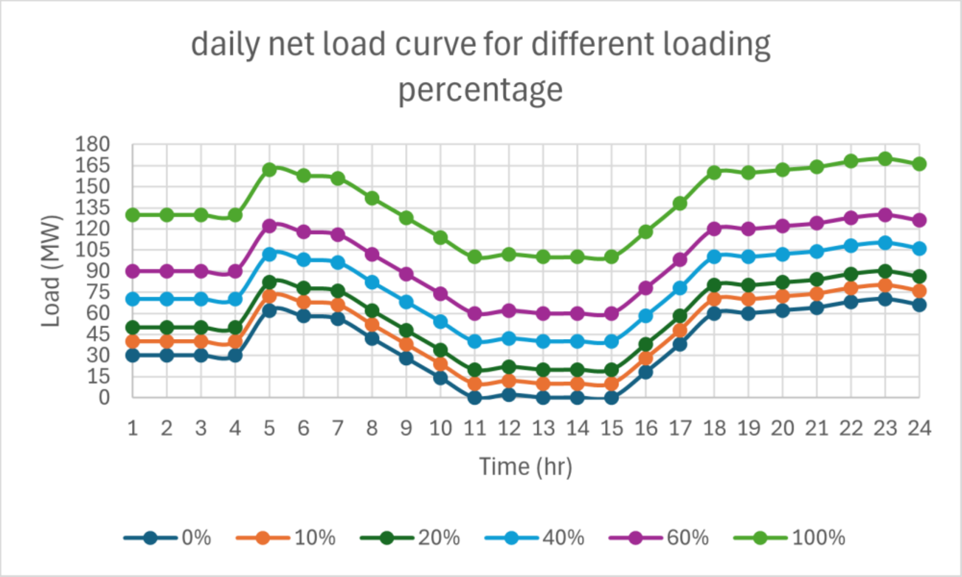


Figure 2.7: Data center daily net load at different percentages of loading

Table 2: Load factor with corresponding loading

Loading	0 %	10 %	20 %	40 %	60 %	100 %
Load Factor	53%	59%	67%	70%	78%	81%

2.5.2 Environmental Impacts

The dramatic growth of data centers leads to both increased power consumption and carbon emissions, and data centers must utilize energy efficiently and flexibly if they want to reduce energy consumption and carbon emissions.

A higher carbon intensity denotes a larger environmental impact, whereas a lower carbon intensity points to a cleaner energy mix with fewer greenhouse gas emissions per unit of electricity generated. The energy used is renewable in Norway and Nordic countries, in contrast to countries such as Germany and the UK where the CO2 intensity per kWh (gCO_2eq/kWh) is up to twelve times as high, according to Electricity Maps. The overview shows that Norwegian consumption consists of 96 percent renewable energy and that the emission is 36 grams of CO2 equivalent per kWh. This emission is primarily due to the use of gas in the traditional powerintensive industry. Texas emits 1.5 metric tons of CO2 per megawatt hour of electricity generated.[2, 23]

The calculation of the daily energy emission with different loading in three different countries can be found in Table 3, and a visual comparison is shown in the Figure 2.8.

Table 3: Emission (metric tons) with corresponding loading in three countries

Countries\Loading	0%	10%	20%	40%	60%	100%
Norway	31.968	40.608	49.248	66.528	83.808	118.368
Germany	373.848	474.888	575.928	778.008	980.088	1384.248
Texas	1332	1692	2052	2772	3492	4932

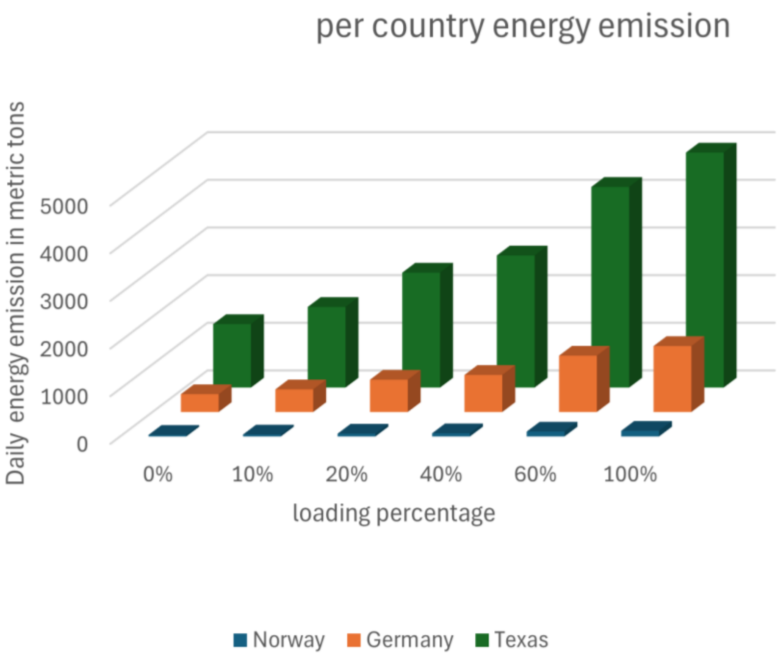


Figure 2.8: Daily energy emission in metric tons in three countries for different percentages of loading

2.5.3 Cost Estimation

The power supply system, networking infrastructure, and cooling resources are the three fundamental building blocks of a data center's operation [24]. The cost calculation is done on paper [25] exclusively assumed to be focused on the construction and use phases of the data center and ignores the cost of facility decommissioning and materials and equipment disposal.

In Figure 2.9 below, the site infrastructure capital cost contributes 42 %, IT capital cost is 32 %, other operating expenses are 14 % and Energy costs are 12 % approximately.

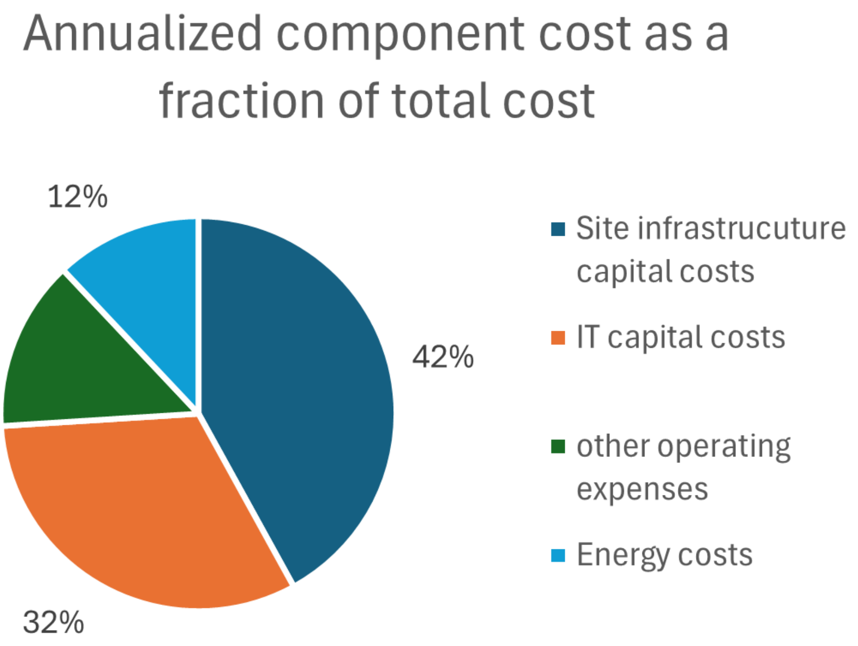


Figure 2.9: Annualized component cost as a fraction of total cost [25]

The site infrastructure and IT capital costs include hardware price (the acquisition cost of IT devices including servers disk, tape storage and networking, the power distribution and cooling equipment, and other operating systems acquisition cost, basic installation and design/engineering cost, land cost, and project management/facility engineering cost.

Other operating costs are power consumption costs (the cost of electricity for servers, networking equipment, and cooling), personal costs (staff salaries), maintenance and repairs costs, depreciation costs(amortization), and IT software license costs.

Total operating expenses include electricity costs, network fees, and other operating expenses. The other operating expenses are IT site and facilities site management staff, maintenance, security costs, property taxes, and so on.

Total electricity use for both IT and non-IT loads is the sum of electricity consumed by the IT, cooling, and auxiliaries. Later two fall into the non-IT load. Direct IT power is the product of watts per rack times the number of racks present and non-IT loads are cooling electricity consumption includes chillers, fans, pumps, CRAC units, and auxiliaries including UPS/PDU losses, lights, and other losses.

The cost estimation done by Total cost of ownership (TCO) models depends on the size, location, and design of the data center and includes capital costs and other operational costs. Tier 4 data centers can cost USD 22 million/MW in comparison to USD 10 million/MW for a ‘Tier 1’ facility [26]. Taking three-quarters of capital cost (CapEx) for both infrastructure and IT and one-quarter of the cost to operating cost (OpEx) that includes the other operating costs and energy costs.

Considering that the energy cost increases linearly with the total energy consumption of the data center, the energy cost can be calculated for different loading percentages. However, energy cost includes not only electricity but any other form of energy used for cooling and other purposes.

Since Texas, 100 MW is the Tier 4 data center the total approximate cost will be USD 2200 million which is 75 % of the total capital cost with the remaining 25 % as shown in Table 4.

Table 4: Approximated total cost of 100 MW tier 4 data center distributed into different costs.

Total capital Cost (75%)		Other operating expenses + Energy Cost (25 %)
Site Infrastructure Capital Cost	IT Capital Cost	
USD 1650 million		USD 550 million

Within data centers, servers and other IT equipment consist of both hardware and software or virtual components. Total cost of ownership (TCO) models were proposed by most data center TCO models, but they only included the cost of the hardware of the IT devices and ignored the costs of their other virtual components. The findings in the paper indicate that the cost associated with software can be greater than those of the hardware in IT products. Therefore, when constructing a data center, it is important to account for both costs to calculate the total costs of the data center. Taking a total of 10 servers, the hardware price of server machines was found to be 43,680 $ but the total software license cost was 6,933,456$ [26] which shows that the software license cost can be greater than the hardware cost. Hence, while calculating the overall cost of the data center both should be taken into consideration.

3 Modeling and Simulation

3.1 Optimization Model

An essential component of many systems, including cloud systems and power grids is the efficient and optimal serving of various types of workloads within a given time frame. Load balancing techniques and scheduling algorithms play a crucial role in serving the various types of workloads efficiently and optimally. Client quality of Service (QOS) is the responsibility of the cloud service provider. Cloud service providers deal with a variety of challenges as the number of client requests in the cloud environment rises, including resource scheduling and allocation, security, privacy, and virtual machine migration.

In a data center, the amount of resources required for a given resource type: computing (CPU usage), memory usage, task duration, and disk usage is known as workload. Datacenter workload classification and characterization are critical in workload analysis and its accuracy is equally important as well which contributes to the prediction and allocation of the resources in the data center. In the paper [27], the data center workloads are classified and characterized based on resource usage which is critical for capacity planning, resource provisioning, and task scheduling. It has been identified that in the Google Cluster Trace (GCT) dataset, task duration is the key factor with 75.82 % domination that differentiates different workloads from each other. These workloads of cloud data centers are more heterogeneous and non-linear. Heterogenous means the varieties of tasks or processes differing in the computational requirement, resource usage, or data characteristics, and non-linear means that even if the resources are doubled or increased the performance may or may not improve. The detailed characteristics of each workload type can be found in the Table 5 below.

Table 5: Four types of workloads in Google Cluster Trace (GCT) [27]

Workload Type	Duration	Characteristics	Example
A	Short-duration tasks in a large number	Low CPU, memory, and disk usage	Consumes a large amount of resources. An example is small queries to the application
B	Long duration with medium CPU and memory	Medium CPU and memory and low disk usage	An example is a background task running for an application
C	Long duration with low CPU and memory	Long duration but low resource usage for CPU, memory, and Disk	Alive but less active. An example is comprised of a demon process that runs in the background, logging and backup tasks and monitoring
D	Long duration but high server	Long duration with high CPU and memory but low disk usage	Examples are scheduler and core handlers

To cost-effectively meet the application's performance requirements, data center workload modeling is required where prioritizing the different workloads impacts the total profit made by the data center. It has been observed that the user application is not using as many resources as it has requested.

Model Description based on Profit Maximization Problem:

In this section, we first present the problem formulation of the optimization objective of the model. We then introduce the hourly load curve, revenue, and energy cost values for workload types which are the synthetic data derived from the real-world data.

Problem Formulation:

We first introduce the notation ‘i’ types of workloads based on the four characteristics namely CPU usage, memory usage, task duration, and disk usage. Given workload type ‘i’, the optimization goal is to choose a workload as per prioritization each hour in such a way that the profit generated by the data center is maximum and the total demand in a day doesn’t exceed the capacity of the data center. The total profit of the data center at any hour is calculated as:

Total Profit (P) = Revenue (R) – Cost (C)

The optimization problem is expressed as follows:

Problem:

Maximization: $P\colon \sum_{h=1}^{24}(\sum_{i=1}^{4} R_i(t) - \ C_i(\mathrm{t})).X_i(t)$
where,

P (EUR) is the total profit generated by serving all types of workloads over 24 hours

R_i(t) is the hourly revenue generated by serving the workload i.

C_i (t) is the hourly cost by serving the workload type i.

$X_i(t)$ is the binary decision variable indicating whether to serve workload i at time *t* and is dimensionless.

Decision variables: $X_i(t)$

- $X_i(t) = 1,$ if workload i is served at time t.
- $X_i(t)$ = 0, otherwise.

Subjected to:

Demand constraint ensuring total demand does not exceed data center capacity:

$\sum_{i=1}^{4}\sum_{h=1}^{24}\ X_i(t).\,W_i \leq$ total capacity of the data center where $W_i(t)$ is the load demand per hour for each workload.

Workload Prioritization Constraint:

Prioritize workload based on weighted value $(V_i(t))$

$\sum_{\boldsymbol{t=1}}^{\boldsymbol{24}} X_1(t)\ .V_1(t) \geq \sum_{\boldsymbol{t=1}}^{\boldsymbol{24}} X_2(t).\,V_2(t) \geq \sum_{\boldsymbol{t=1}}^{\boldsymbol{24}} X_3(t).\,V_3(t) \geq \sum_{\boldsymbol{t=1}}^{\boldsymbol{24}} X_4(t).\,V_4(t)$

Total profit (P) ≥ 0

Hourly Load Curve of Workloads:

Considering the analyzed GCT dataset's workload types led to the assumption that a similar pattern exists, with the total incoming workload each hour consisting of a combination of these four workloads, with workload A being more dominant than workload B, and so on. As a result, the percentages of workload A, B, C, and D represent 60%, 22%, 14%, and 4% of the total incoming load per hour, respectively. The hourly load for these four types of workload can be found in Appendix B and is visualized in the Figure 3.1.

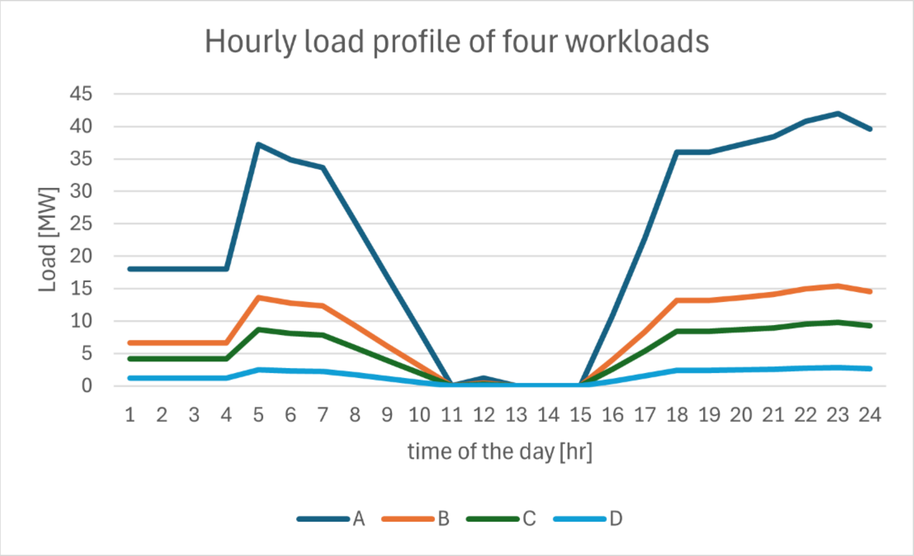


Figure 3.1: Hourly load profile of four workload types

Energy Cost and Revenue Generation Curve:

The electricity price trace from the Nordpool website for Region 1[2] was used to generate the synthetic set of electricity prices and uses the electricity price as a reference for revenue generation assumption to provide input on the electricity price and revenue generated for four workloads. Appendix C contains all of the energy costs and revenue generated for the four workloads, which are displayed in the Figure 3.2 and Figure 3.3.

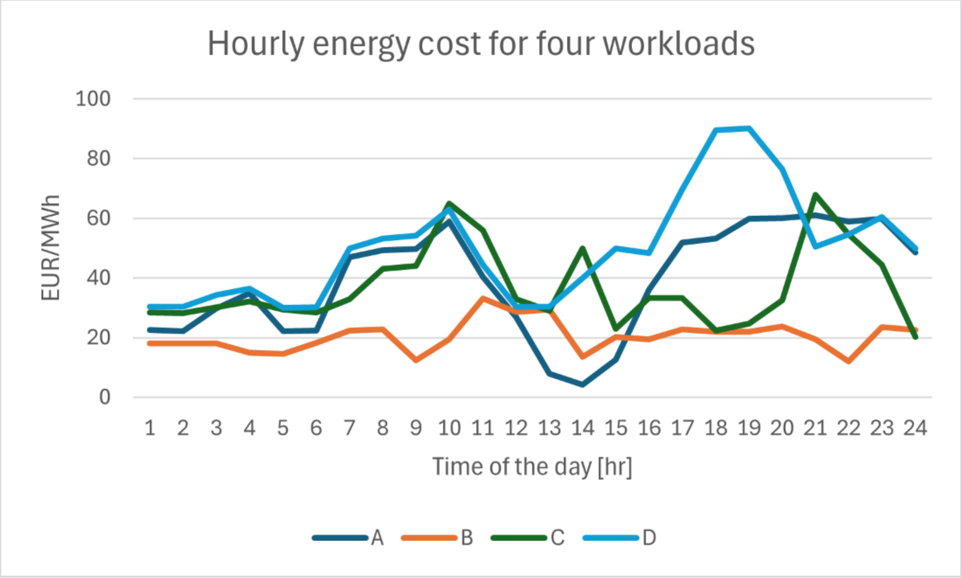


Figure 3.2: Hourly energy cost based on Region 1 in Norway for four workload types

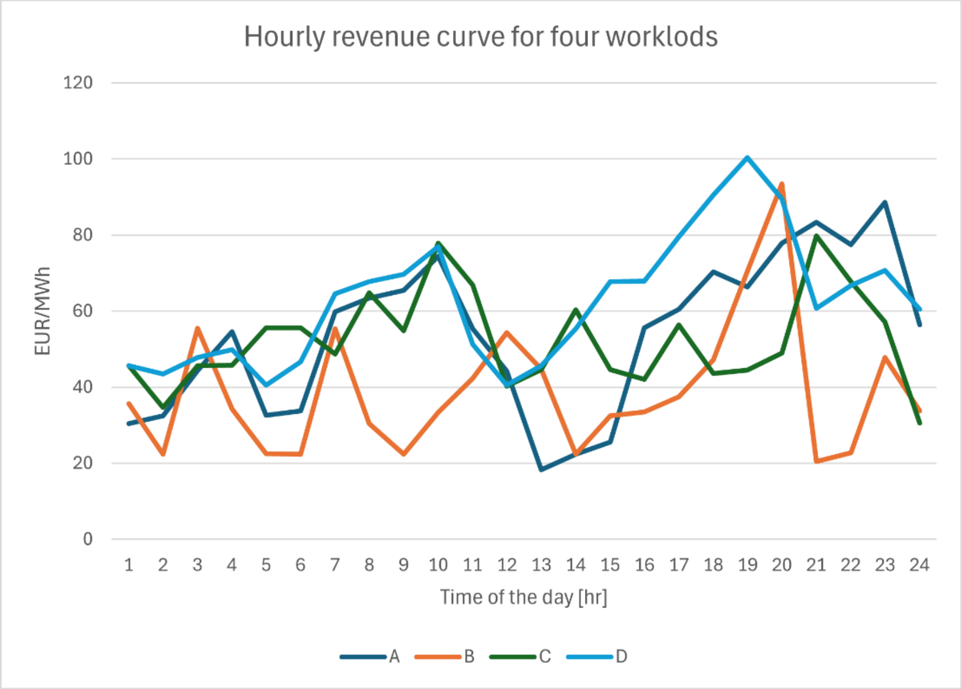


Figure 3.3: Hourly revenue generation curve for four workload types

3.2 Simulation model

DIgSILENT PowerFactory simulations:

The simulations were conducted in DIgSILENT PowerFactory to investigate the impact on the IEEE Nine-bus system under the different loading of the data center, with a specific focus on total daily grid loss.

Introduction to DIgSILENT PowerFactory :

PowerFactory is one of the best power system analysis tools for industrial, distribution, transmission, and generation systems. The full spectrum of functionality is covered, ranging from fundamental functions to extremely complex and advanced applications, such as distributed generation, wind power, real-time simulation, and performance monitoring for system oversight and testing. Modern algorithms, a distinctive database design, and dependable and adaptable system modeling capabilities are all combined in PowerFactory's user-friendly, Windows-compatible interface. Additionally, PowerFactory is ideal for highly automated and integrated solutions in business applications also due to its flexibility in scripting and interfacing.[28]

For medium to long-term simulation studies, PowerFactory comes with the Quasi-Dynamic Simulation toolbox, a specialized time-varying load flow calculation tool. With the ability to choose the simulation period and step size, this tool performs a sequence of load flow simulations that are spaced in time.

Since the majority of operational parameters depend on time. For example, the load is dependent on time due to daily and seasonal cyclic load variation and renewable energy sources like solar and wind power vary according to solar insolation and wind speed, which is again the function of time itself. A sensible and practical method is to use a sequence of load-flow calculations with different model parameters that are time-dependent to simulate so-called "Quasi-Dynamic" phenomena.

For this report simulation, Qausi- Dynamic Simulation (time-varying load flow analysis) is done to calculate the losses in the grid under replacing constant loads at different nodes with our 100 MW data center time-varying load profile generated in chapter 2.5.1 under mainly three different percentages of loading.

Introduction to IEEE Nine-bus system:

P. M. Anderson and A. A. Fouad's book Power System Control and Stability introduced the Nine-bus System. The Figure 3.4 represents the single-line diagram of the Nine-bus transmission system, which is made up of nine buses (nodes), three generators, three loads, six lines, and three transformers as built in the power factor.

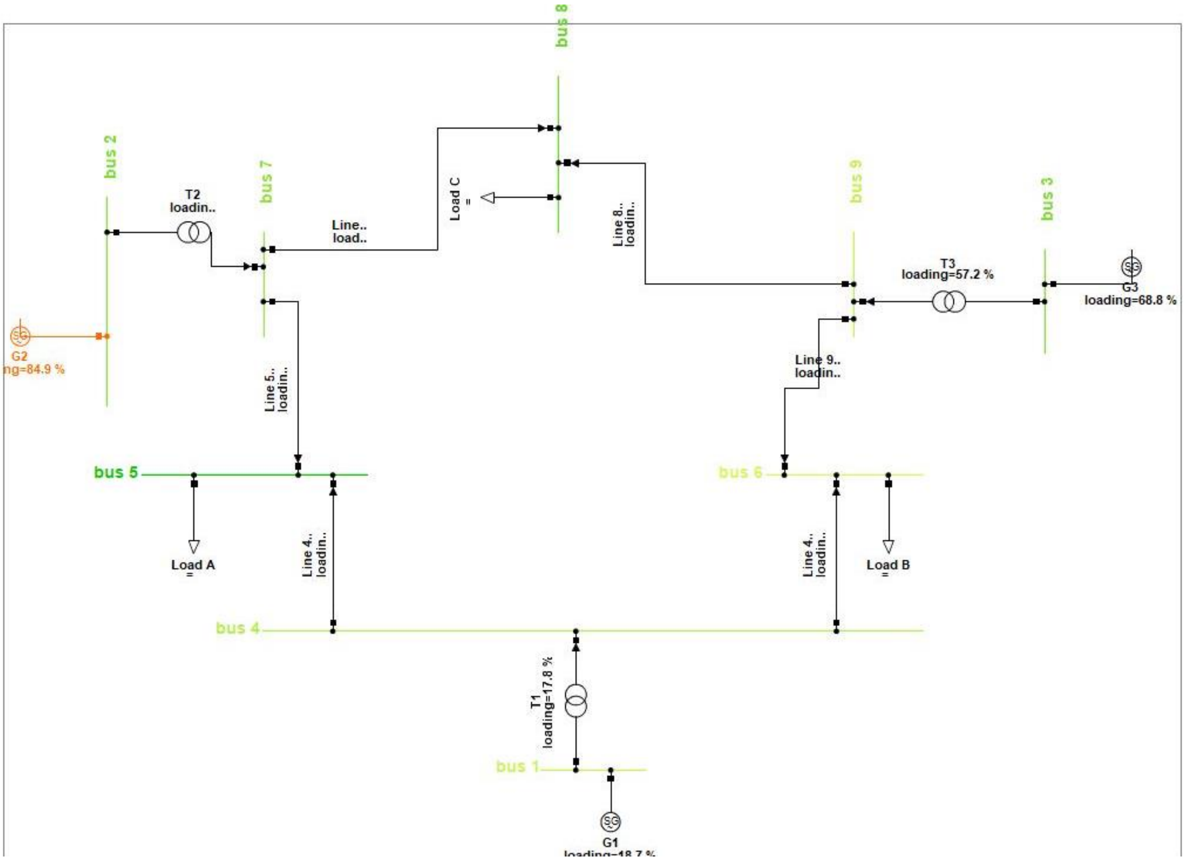


Figure 3.4: Single line diagram of the IEEE Nine-bus system built in DIgSILENT

Table 6: Constant loads

Load	Bus	P [MW]	Q [Mvar]
Load A	Bus 5	125	50
Load B	Bus 6	90	30
Load C	Bus 8	100	35

3.3 Scenarios and Results

3.3.1 Result from optimization model

The prioritized load curve was created using the serving decision variables that the solver provided after running the Excel model for the profit maximization problem. The value of the prioritized hourly load curve each hour can be seen in Appendix C and is displayed in the Figure 3.5 which shows that the total daily demand was 99.60 MWh with a peak demand of 16 MW at 20:00 hour.

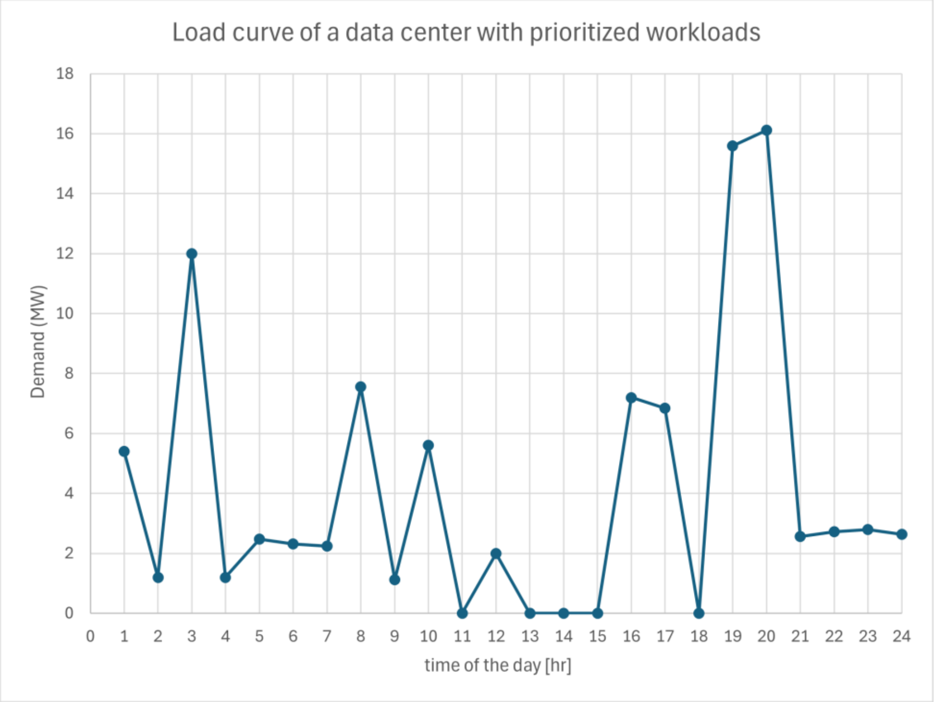


Figure 3.5: Hourly load curve with the prioritized workloads

The feature of this model is that the workloads are served based on the required prioritization set by the data center or demanded by the enterprises and have the flexibility of changing the priority as per the requirement.

3.3.2 Scenario Evaluation

In the test scenario, the prioritization level was changed, and the model was tested for different prioritization levels. Table 7 shows that the highest total profit resulted when workload type A was prioritized first, followed by prioritization levels A>B>C>D.

Table 7: Test scenario with different prioritization levels and corresponding total profit generation

Prioritization level	Highest Prioritized Workload	Total daily demand with prioritized workload (MWh)	Profit (EUR/MWh)	Total Profit (EUR)
A>B>C>D	A	99.60	795.03	79184.99
B>A>C>D	B	99.60	779.94	77682.02
C>A>B>D	C	100.00	774.23	77423.00
D>A>B>C	D	100.00	774.23	77423.00

The solver initially was unsuccessful in converging due to the pre-set constraints limitation, so the prioritization constraint was manually changed, resulting in the value in the model. Even though changing the priority level does not significantly affect the overall profit, the type of workloads that are served each hour varies depending on the assigned level of prioritization. However, all the workloads might not be served each hour.

3.3.3 Impact of load profiles of data centers on the IEEE Nine-bus Power System

This chapter includes the impact analysis of load profiles of data centers on the IEEE nine-bus system. Performing load flow analysis on the original IEEE Nine-bus resulted in the expected grid loss of 4.62 MW under the original constant load. The three scenarios where the time-varying data center loads, replace the original constant loads at the three buses at loading percentages of 0%, 40%, and 100% were studied as three different cases.

The total grid loss curve for each case under three loadings can be found in Appendix D. Reactive power for data center loads is set to 0 Mvar since they do not consume any reactive power.

Case 1: Placement of data center at bus 6

The data center 24-hour load replaced load B at bus 6, and the corresponding total grid loss is presented in the Table 8 at various loading percentages.

Table 8: Loading percentage and respective total daily grid loss, maximum and minimum loss in case 1

Loading percentage	Total daily Grid loss (MW)	Minimum and maximum loss (MW)	Total Generation (MW)	Total Load (MW)

0 %	4.52	4.51 – 4.97	293.52	289
40 %	4.80	4.52 - 4.88	333.80	329
100 %	6.10	4.75 – 6.30	395.10	389

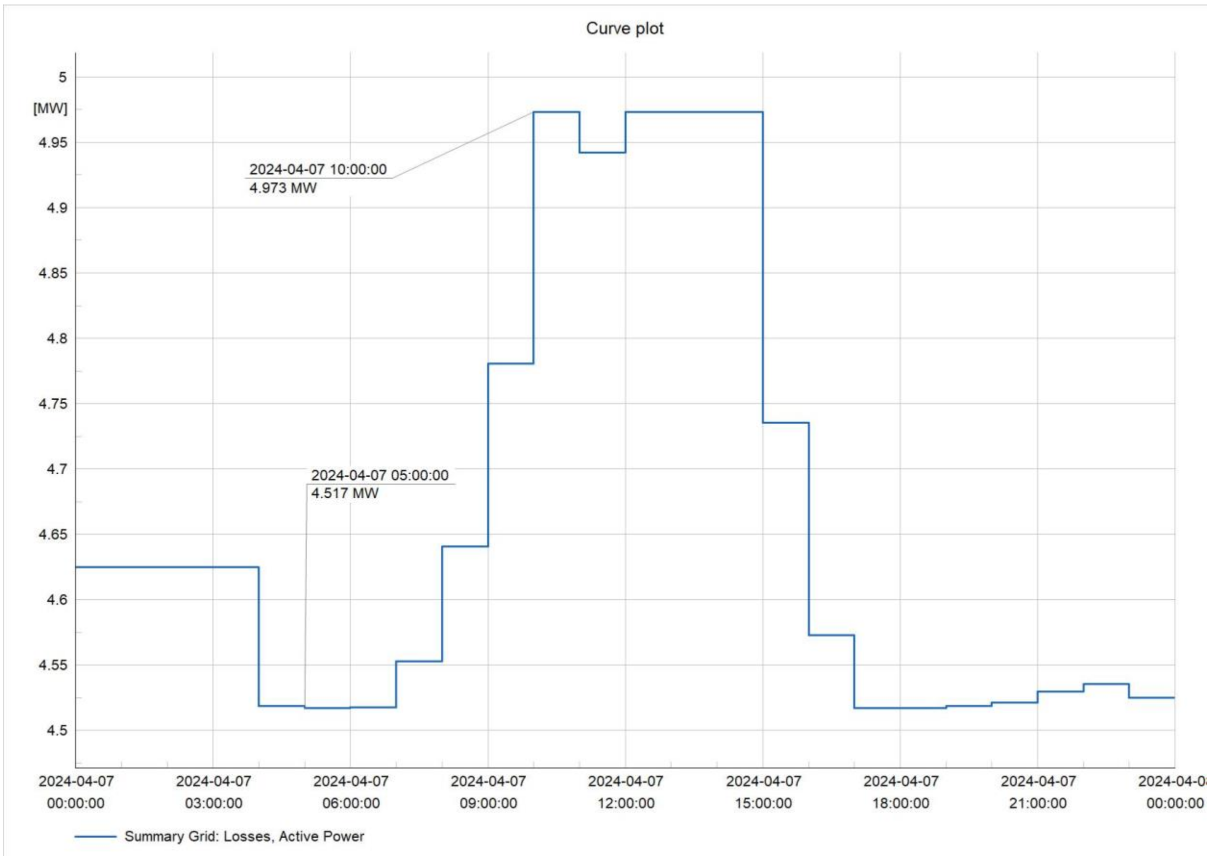


Figure 3.6: Total grid losses (active power) at 0% loading showing the time of maximum and minimum loss

Case 2: Placement of data center at bus 5

The data center 24-hour load replaced load A at bus 5, and the corresponding total grid loss is presented in the Table 9 at various loading percentages.

Table 9: Loading percentage and respective total daily grid loss, maximum and minimum loss in case 2

Loading	Total daily Grid loss (MW)	Minimum and maximum loss (MW)	Total Generation (MW)	Total Load (MW)
0 %	4.37	4.355 - 4.907	258.37	254
40 %	4.39	4.343 - 4.49	298.39	294
100 %	4.94	4.372 - 5.03	358.94	354

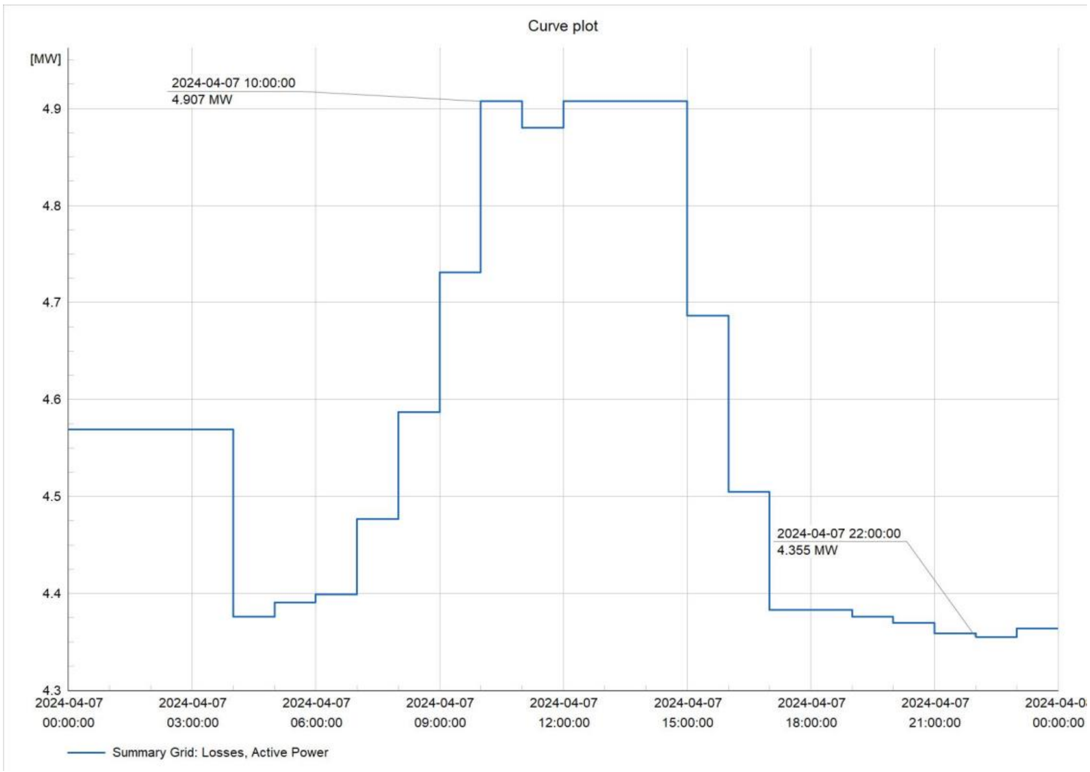


Figure 3.7: Total grid losses (active power) at 0% loading showing the time of maximum and minimum loss

Case 3: Placement of data center at bus 8

The data center 24-hour load replaced load C at bus 8, and the corresponding total grid loss is presented in the Table 10 at various loading percentages.

Table 10: Loading percentage and respective total daily grid loss, maximum and minimum loss in case 3

Loading percentage	Total daily Grid loss (MW)	Minimum and maximum loss at different times in a day (MW)	Total Generation (MW)	Total Load(MW)
0 %	6.05	5.74 – 10.55	285.05	279
40 %	4.41	4.24 – 7.46	323.41	319
100 %	3.66	3.64 – 4.53	382.66	379

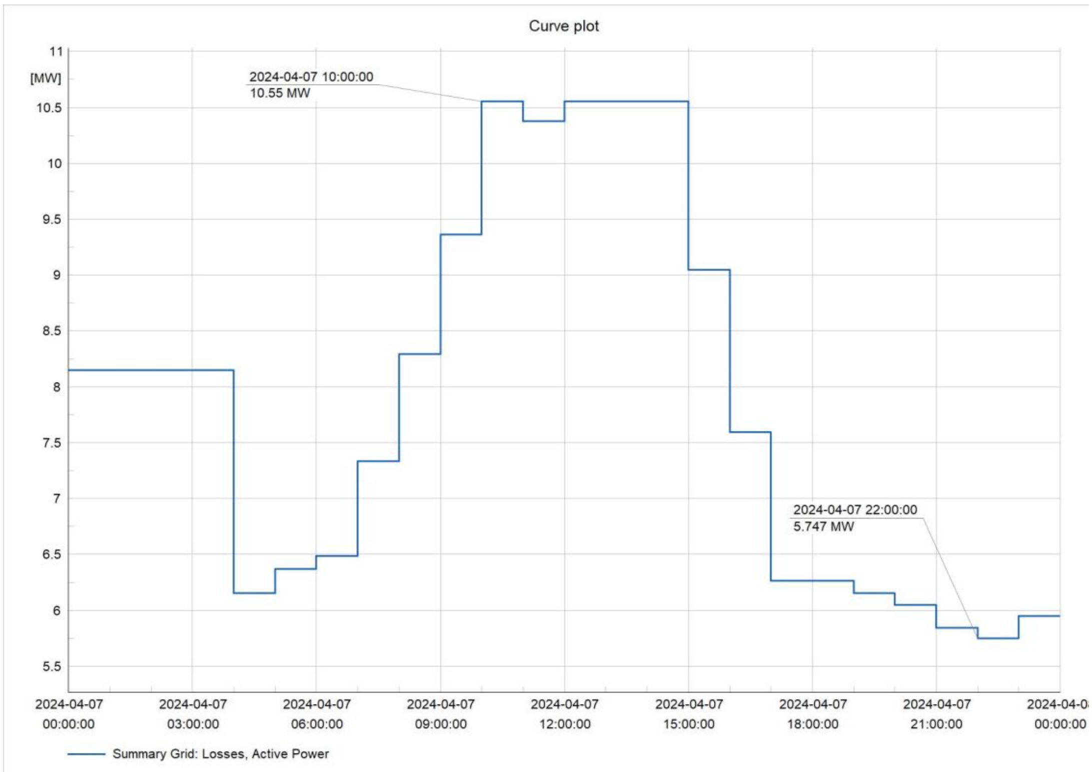


Figure 3.8: Total grid losses (active power) at 0% loading showing the time of maximum and minimum loss

3.3.4 Impact of load curve with prioritized workload on the IEEE Nine-bus Power System

The detailed simulation result curve of a load curve with the prioritized level A>B>C>D for each three cases with the time of maximum and minimum loss can be seen in Appendix D and are summarized in the Table 11 below.

Table 11: Total daily grid loss with load curve with prioritized workloads in three cases

Case	Grid loss (MW)	Maximum and minimum loss (MW)
1	4.93	4.97 - 4.75
2	4.87	4.90 - 4.70
3	10.32	10.55 - 9.91

In case 2, the total daily grid loss is lower than in other cases, making it feasible and efficient to place the data center with prioritized workloads on bus 5.

3.4 Discussion

The impact analysis of data center load profiles on the IEEE nine-bus system under three scenarios or cases for different percentages of loading showed that placing the data center on bus 8 in case 3 appears to be the most reliable and optimum due to the decrease in total loss with increasing loading, as opposed to the other two cases. The total grid loss at 100% loading was found to be 3.66 MW, a 20.77% reduction from the initial total grid loss of 4.62 MW.

The impact analysis of integrating the prioritized load curve with the prioritized level A>B>C>D shows that placing the data center on bus 5 in case 2 is more feasible and efficient, resulting in a total daily grid loss of 4.87 MW less than the other two cases. This total grid loss is slightly higher than the original 4.62 MW, but this difference can be caused by several factors, including total load demand, load distribution, network configuration, line impedance, and operating conditions. A detailed power flow analysis, including an examination of the specific paths of power flow and the condition of the network infrastructure, would be required to identify the exact causes, which could be a future scope of research.

4 Conclusion

4.1 Main Conclusion

To understand the data center operation and its types, a comparative study of data centers based on the scale and performance was carried out. It also presents how the data center's underutilized backup resources can be used as batteries and that can contribute to the overall efficiency and sustainability of the power system grid and potentially provide financial benefits to the data center operators.

Some of the widely implemented load balancing strategies are Time shifting, Geography shifting, adjusting power load, and UPS energy storage systems that are discussed in detail and make the data centers more flexible. Demand-side management (DSM) strategies are emphasized, requiring load flexibility and data transfer capabilities. The load curve of the data center under various loading has been presented and the corresponding load factor also known as capacity factor provided the availability of the energy resources in the data center that can participate in the energy optimization strategies of the data center. The environmental impacts with the corresponding loading were quantified showing Norway's suitability for environmentally sustainable and efficient data center operations A simple explanation of the annualized cost of the data center was presented as well.

The maximum profit was obtained by integrating the workload prioritization strategy with prioritization level A>B>C>D, hourly load curve, electricity cost, and revenue generation into the Excel model. Prioritizing workloads based on priority, electricity costs, and potential revenue generation ensures that critical tasks are completed efficiently while remaining cost-effective. and the most feasible location for this data center was investigated through load flow analysis and quasi-dynamic simulation in DIgSILENT PowerFactory. Once high-loss areas are identified, it helps in implementing the necessary optimization strategies and actions to improve grid operations by analyzing grid losses. By analyzing the impact of adding new data center loads on the grid, informed decisions about grid expansion and reinforcement can also be made.

This thesis has made a substantial contribution to the field of data center operations and power system integration, providing a robust framework for future research and practical implementation. The adoption of these methodologies promises to enhance the sustainability, efficiency, and resilience of data centers and power grids alike. The implications of these findings are significant for both data center operators and power system managers. By adopting the methodologies and strategies developed in this thesis, stakeholders can achieve a more sustainable and cost-effective operation, contributing to broader environmental and economic goals.

4.2 Future Work

Future research could focus on improving the proposed optimization problems and strategies, as well as investigating their application in various geographical and operational contexts and incorporating cutting-edge technologies such as artificial intelligence and machine learning to improve operational efficiency even more. Contingency analysis can be used to conduct further studies on overall grid reliability in the context of data center flexibility.

One of the aims of this thesis is to help balance energy consumption on the power grid, contributing to grid stability and reliability which could done through contingency analysis in the future.

References

[1] M. E. a. D. Consulting, "How to establish a data center in Norway." Accessed: 1 march 2024. [Online]. Available: https://www.regjeringen.no/contentassets/a76ebef545ae4e87a5b0761a93fb6ba1/how-to-establish-a-data-center-in-norway.pdf

[2] N. D. c. Industry, "The Data Center Industry in Norway 2023-2024," 2023.

[3] F. G. o. Gül Nihal Gü˘gül, Ursula Eicker, "Sustainability analysis of zero energy consumption data centers with free cooling, waste heat reuse and renewable energy systems: A feasibility study," *ScienceDirect,* vol. 262, 2023.

[4] H. Z. Huigui Rong, Sheng Xiao, "Optimizing energy consumption for data centers," *ScienceDirect,* pp. 674-691, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032115016664?via%3Dihub.

[5] m. Company. "Investing in the rising data center economy." https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy (accessed 3 february, 2024).

[6] Y. W. Yanwei Zhang, and Xiaorui Wang, "GreenWare: Greening Cloud-Scale Data Centers to Maximize the Use of Renewable Energy," 2011.

[7] R. Urgaonkar, "Optimal Power Cost Management Using Stored Energy in Data Centers," 2011.

[8] R. W. W. C. L. L. a. K. ZHAO, "Energy efficient optimization method for green data center based on cloud computing," 2015.

[9] M. Osman. "Beginner’s guide to different types of data centers." https://www.nexcess.net/blog/types-of-data-centers/ (accessed 12 March, 2024).

[10] K. Hitchens. "Understanding the Differences Between 5 Common Types of Data Centers." https://www.datacenterfrontier.com/sponsored/article/11427373/belden-understanding-the-differences-between-5-common-types-of-data-centers (accessed 13 March, 2024).

[11] S. E. L. H.S. Sun "Case study of data centers’ energy performance," *Elsevier,* vol. 38, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778805001738#bib1.

[12] u. Institute. "Tier Classification System." https://uptimeinstitute.com/tiers (accessed 21 march 2024).

[13] S.-G. Cooperation. "Greener Data Center: Flexibility and Renewables." Energiepartnerschaft https://www.energypartnership.cn/home/greener-data-center-flexibility-and-renewables/ (accessed 2 April 2024).

[14] P. T. L. B. T. M. H. R. F. U. Tamrakar, "Real-time Operation of a Data Center as Virtual Power Plant Considering Battery Lifetime," 2018 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8445345/authors#authors

[15] D. Sidhu, "Data Center Battery Systems - Virtual Power Plants," ed, 2023.

[16] S. R. A. S. C. R. Tonkoski, "Operation of Datacenter As Virtual Power Plant " *IEEE,* 2015.

[17] S. A. Labi Bajracharya, Santosh Chalise, Timothy M. Hansen, and Reinaldo Tonkoski, "Economic Analysis of a Data Center Virtual Power Plant Participating in Demand Response," *IEEE,* 2016.

[18] I. A. S. H. J. Paananen, "Data centers as a source of dynamic flexibility in smart girds," vol. 229, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261918310845#b0220.

[19] P. G. I. Services. "Data Center Tiers Explaine." https://phoenixnap.com/blog/data-center-tiers-classification (accessed 21 March, 2024).

[20] N. D. Center, ""Tier 3 vs Tier 4 Data Centers: Understanding the Differences."," ed, 2023.

[21] B. haritas. "Tier 3 vs. Tier 4 Data Centers: Is The Latter Worth The Extra Investment?

." https://cio.economictimes.indiatimes.com/news/data-center/tier-3-vs-tier-4-data-centers-is-the-latter-worth-the-extra-investment/90525717 (accessed 21 March, 2024).

[22] D. TECH. "COMPARING THE DATA CENTERS - PART 1." https://blog.daouidc.com/en-us/blog/comparing-data-centers-1 (accessed 21 March, 2024).

[23] B. E. M. a. J. Thompson, "Texas' Energy Base Drives Climate Concerns as Renewables Expand," 2019. [Online]. Available: https://www.dallasfed.org/~/media/documents/research/swe/2019/swe1903c.pdf

[24] C. D. Patel, "Cost Model for Planning, Development and Operation of a Data Center," *Researchgate,* 2005.

[25] J. Koomey, "A simple Model for DeterminingTrue toal cost of Ownership for Data Centers," UPTIME INSTITUE. Accessed: 29 March 2024.

[26] S. M. a. E. E. Doaa Bliedy, "Cost Model for Establishing a Data Center," *International Journal of Computer Science, Engineering and Applications (IJCSEA),* vol. 8, 2019.

[27] A. G. Virendra Singh Shekhawat, Ashish Thakrar, "Datacenter Workload Classification and Characterization: An Empirical Approach," *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS),* 2018.

[28] DIgSILENT. POWERFACTORY APPLICATIONS [Online] Available: https://www.digsilent.de/en/powerfactory.html

Appendices

Appendix A

Per hour load demand of a 100 MW data center with different percentages of loading:

Time [hr]	Load (MW) [Reference]	10 %	20 %	40 %	60 %	100 %
1	**30**	40	60	70	110	130
2	**30**	40	60	70	110	130
3	**30**	40	60	70	110	130
4	**30**	40	60	70	110	130
5	**62**	72	92	102	142	162
6	**58**	68	88	98	138	158
7	**56**	66	86	96	136	156
8	**42**	52	72	82	122	142
9	**28**	38	58	68	108	128
10	**14**	24	44	54	94	114
11	**0**	10	30	40	80	100
12	**2**	12	32	42	82	102
13	**0**	10	30	40	80	100
14	**0**	10	30	40	80	100
15	**0**	10	30	40	80	100
16	**18**	28	48	58	98	118
17	**38**	48	68	78	118	138
18	**60**	70	90	100	140	160
19	**60**	70	90	100	140	160

20	**62**	72	92	102	142	162
21	**64**	74	94	104	144	164
22	**68**	78	98	108	148	168
23	**70**	80	100	110	150	170
24	**66**	76	96	106	146	166
total load(MWh)	**888**	1128	1608	1848	2808	3288
Average load(MW)	**37**	47	67	77	117	137
Load factor(%)	**53%**	59%	67%	70%	78%	81%

Appendix B

24-hour load, electricity price, and revenue generated for four workload types,

Hourly load for workload type (MW)			
A	**B**	**C**	**D**
18	6.6	4.2	1.2
18	6.6	4.2	1.2
18	6.6	4.2	1.2
18	6.6	4.2	1.2
37.2	13.64	8.68	2.48
34.8	12.76	8.12	2.32
33.6	12.32	7.84	2.24
25.2	9.24	5.88	1.68
16.8	6.16	3.92	1.12
8.4	3.08	1.96	0.56
0	0	0	0
1.2	0.44	0.28	0.08
0	0	0	0
0	0	0	0
0	0	0	0
10.8	3.96	2.52	0.72
22.8	8.36	5.32	1.52
36	13.2	8.4	2.4
36	13.2	8.4	2.4
37.2	13.64	8.68	2.48
38.4	14.08	8.96	2.56
40.8	14.96	9.52	2.72
42	15.4	9.8	2.8
39.6	14.52	9.24	2.64

Hourly electricity price [EUR/MWh] in Norway reason 1 (NO1)				
Time [hr]	**A**	**B**	**C**	**D**
1	22.49	18	28.4	30.45
2	22.2	18	28.3	30.4
3	29.7	18	30.1	34.3
4	34.92	15	32.2	36.4
5	22.2	14.6	29.4	30
6	22.3	18.2	28.4	30.2
7	46.9	22.4	33	50
8	49.4	22.7	43	53.2
9	49.8	12.5	44	54.2
10	58.99	19.5	65	62.9
11	40.35	33.2	56	44.4
12	26.89	28.6	33	30.3
13	7.95	29.4	29	30.3
14	4.25	13.5	50	40
15	12.68	20.3	23	50
16	36.08	19.4	33.4	48.3
17	51.91	22.7	33.4	69.6
18	53.17	21.9	22.4	89.6
19	59.91	22	24.7	90.2
20	60.02	23.8	32.6	76.5
21	61.03	19.4	67.8	50.4
22	58.93	12.1	54.5	54.5
23	59.92	23.5	44.4	60.4
24	48.5	22.5	20.3	50

Revenue per hour for four workload types [EUR/MWh]				
Time [hr]	A	B	C	D
1	30.4	35.6	45.6	45.6
2	32.4	22.3	34.6	43.4
3	44.3	55.5	45.6	47.8
4	54.6	34.3	45.7	49.8
5	32.6	22.5	55.6	40.5
6	33.8	22.3	55.6	46.7
7	59.8	55.4	48.7	64.5
8	63.4	30.4	64.8	67.8
9	65.4	22.4	54.8	69.7
10	74.4	33.4	77.9	76.9
11	55.4	42.3	66.9	51.2
12	44.2	54.3	40.3	40.5
13	18.3	44.7	44.6	45.6
14	22.3	22.4	60.3	55.4
15	25.5	32.4	44.6	67.8
16	55.6	33.5	42.1	67.9
17	60.4	37.5	56.4	79.5
18	70.3	47.2	43.6	90.5
19	66.3	70.4	44.5	100.4
20	77.9	93.4	48.9	89.5
21	83.4	20.5	79.8	60.7
22	77.5	22.8	67.8	66.7
23	88.6	47.8	57.2	70.7
24	56.4	33.8	30.5	60.4

Appendix C

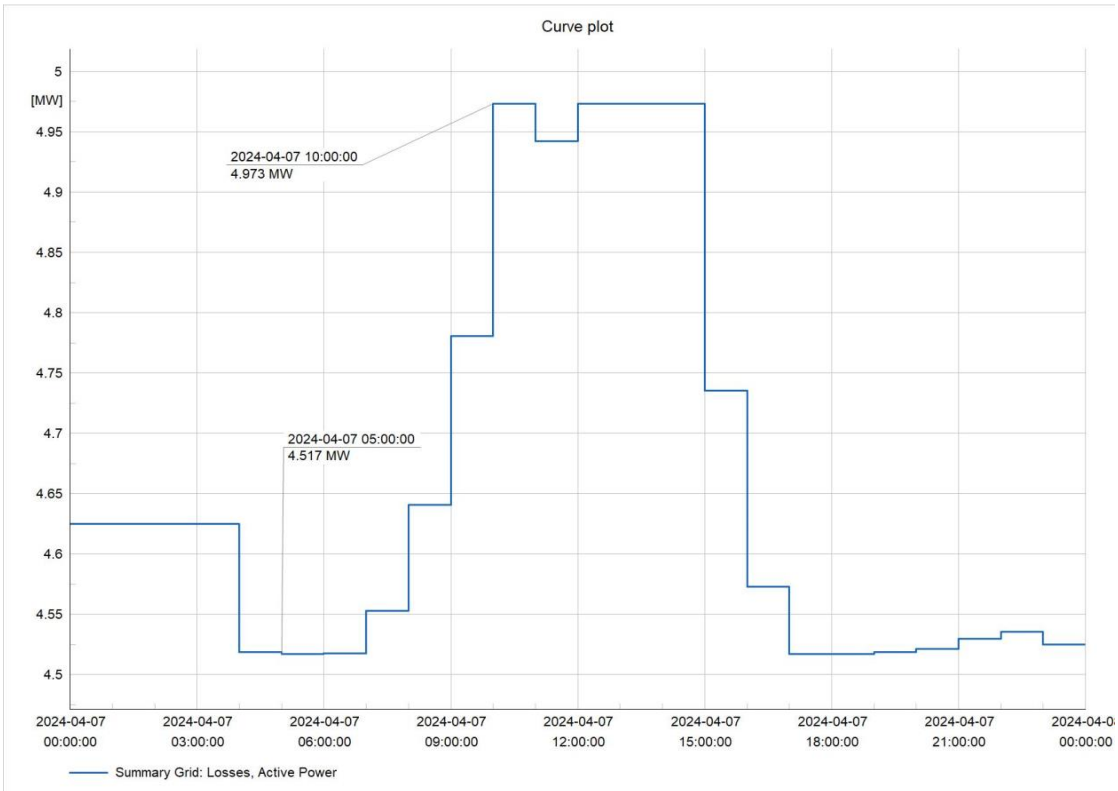
24-hour load data based on the prioritized level A>B>C>D

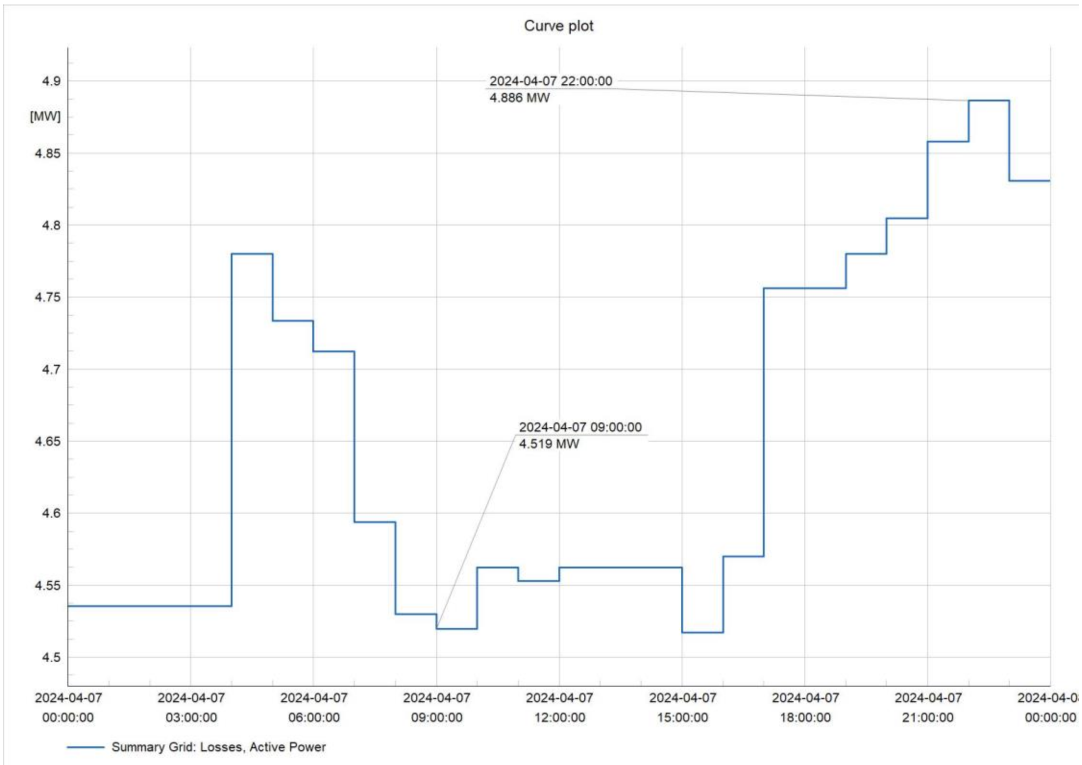
Time (hr)	Serving Decision Variables				Total load (MW)
	A	B	C	D	
1	0	0	1	1	5.4
2	0	0	0	1	1.2
3	0	1	1	1	12
4	0	0	0	1	1.2
5	0	0	0	1	2.48
6	0	0	0	1	2.32
7	0	0	0	1	2.24
8	0	0	1	1	7.56
9	0	0	0	1	1.12
10	0	1	1	1	5.6
11	1	1	1	1	0
12	1	1	1	1	2
13	1	1	1	1	0
14	1	1	1	1	0
15	1	1	1	1	0
16	0	1	1	1	7.2
17	0	0	1	1	6.84
18	0	0	0	0	0
19	0	1	0	1	15.6
20	0	1	0	1	16.12
21	0	0	0	1	2.56
22	0	0	0	1	2.72
23	0	0	0	1	2.8
24	0	0	0	1	2.64

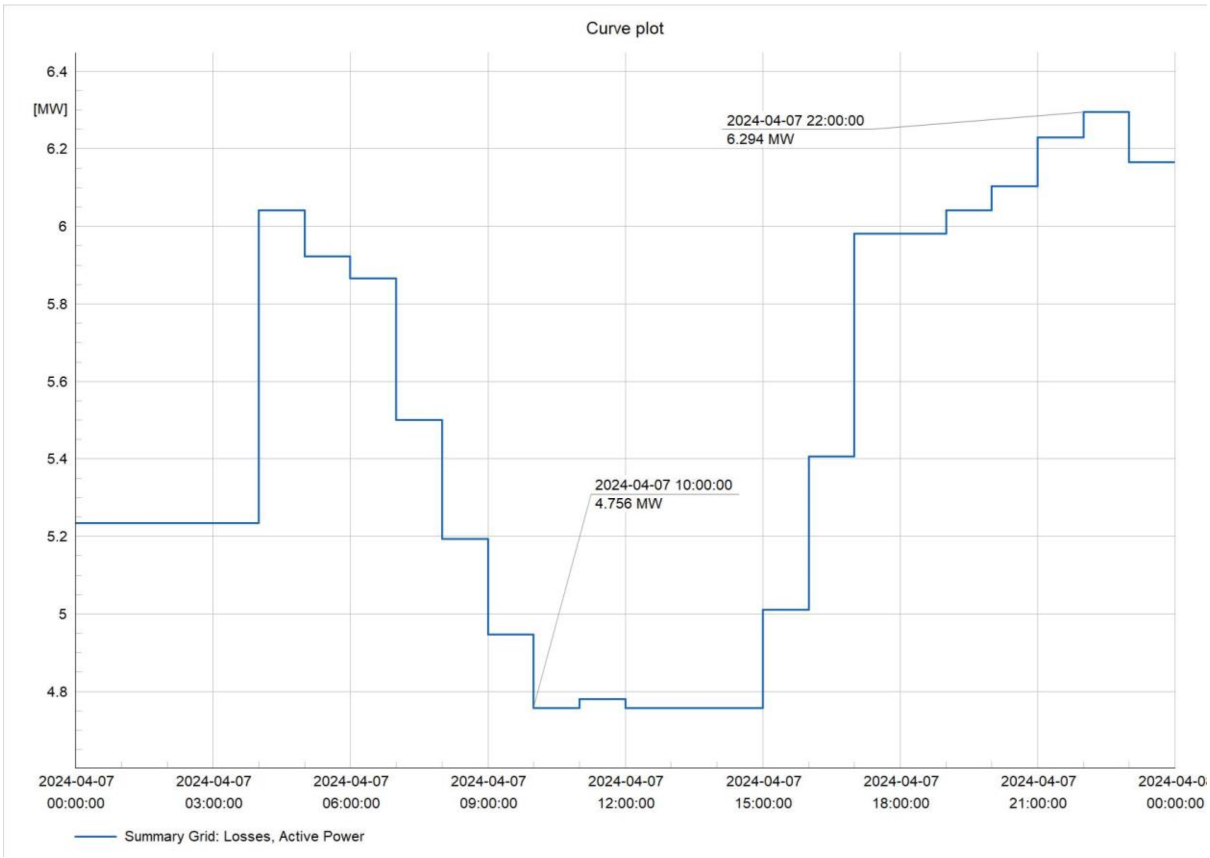
Appendix D

Simulation result curves for 0 %, 40 %, and 100 % loading for three cases

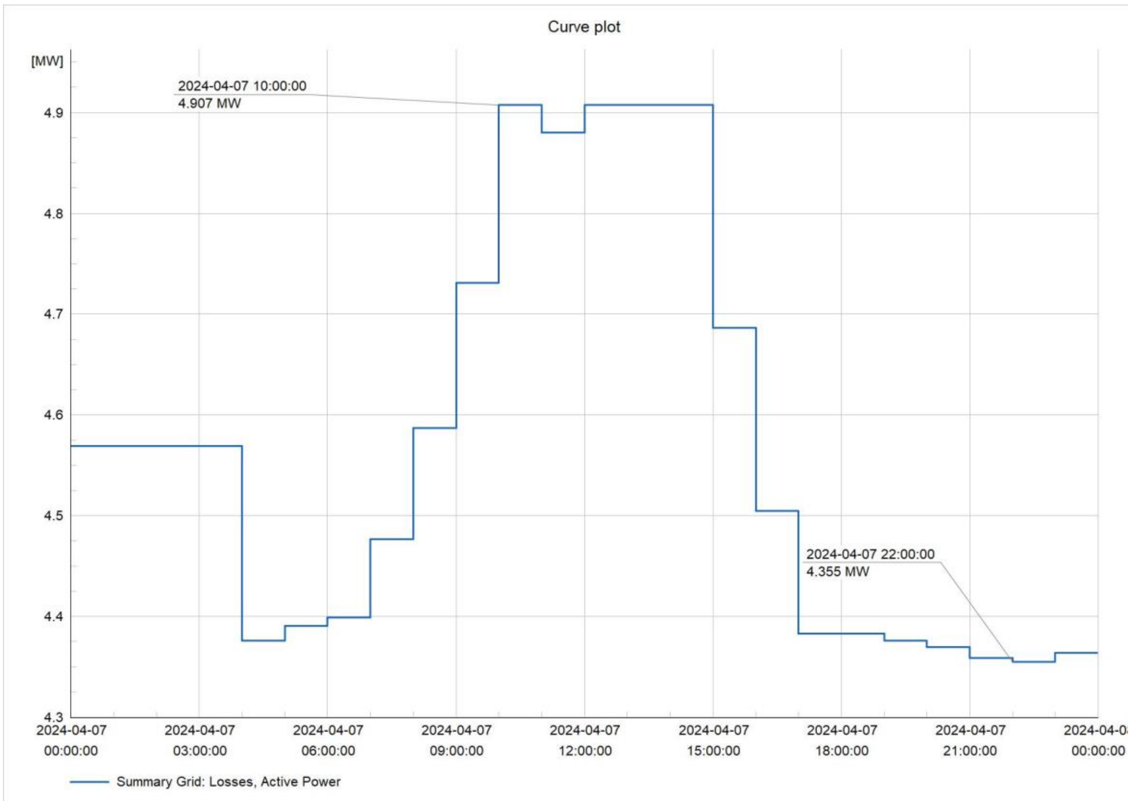
Case 1:

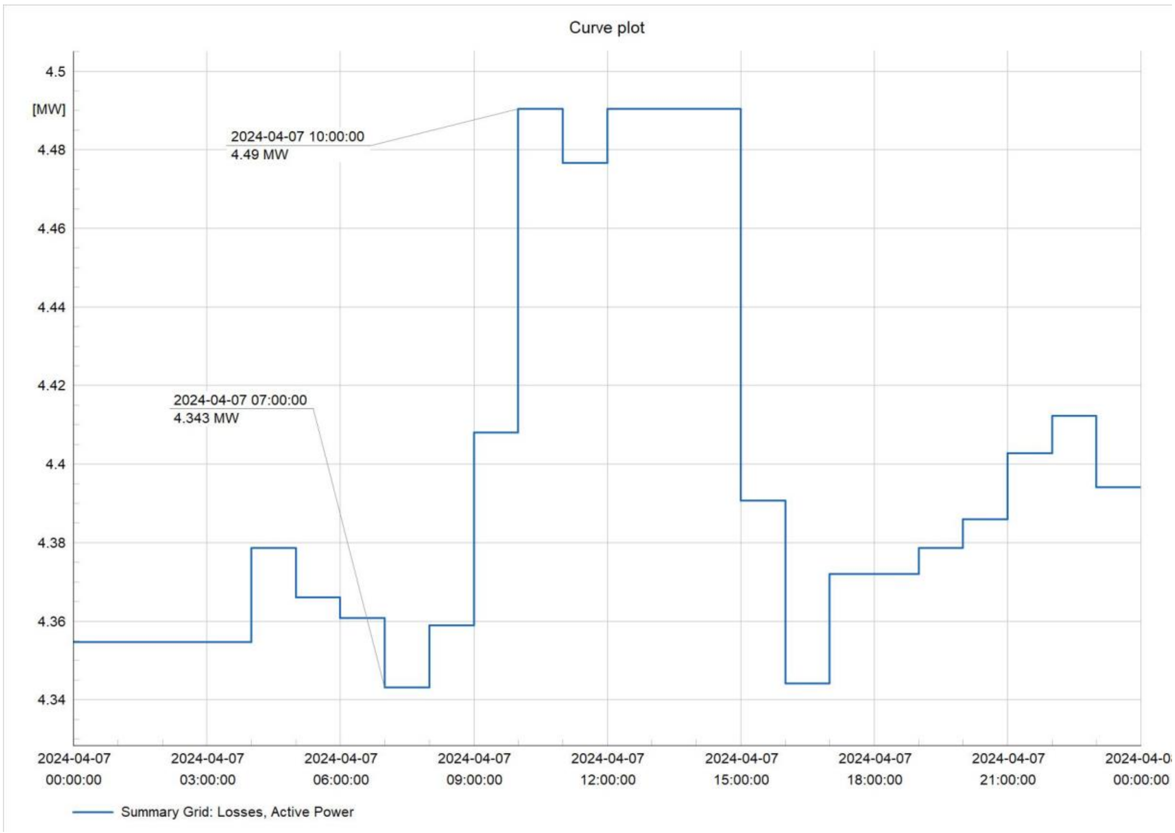


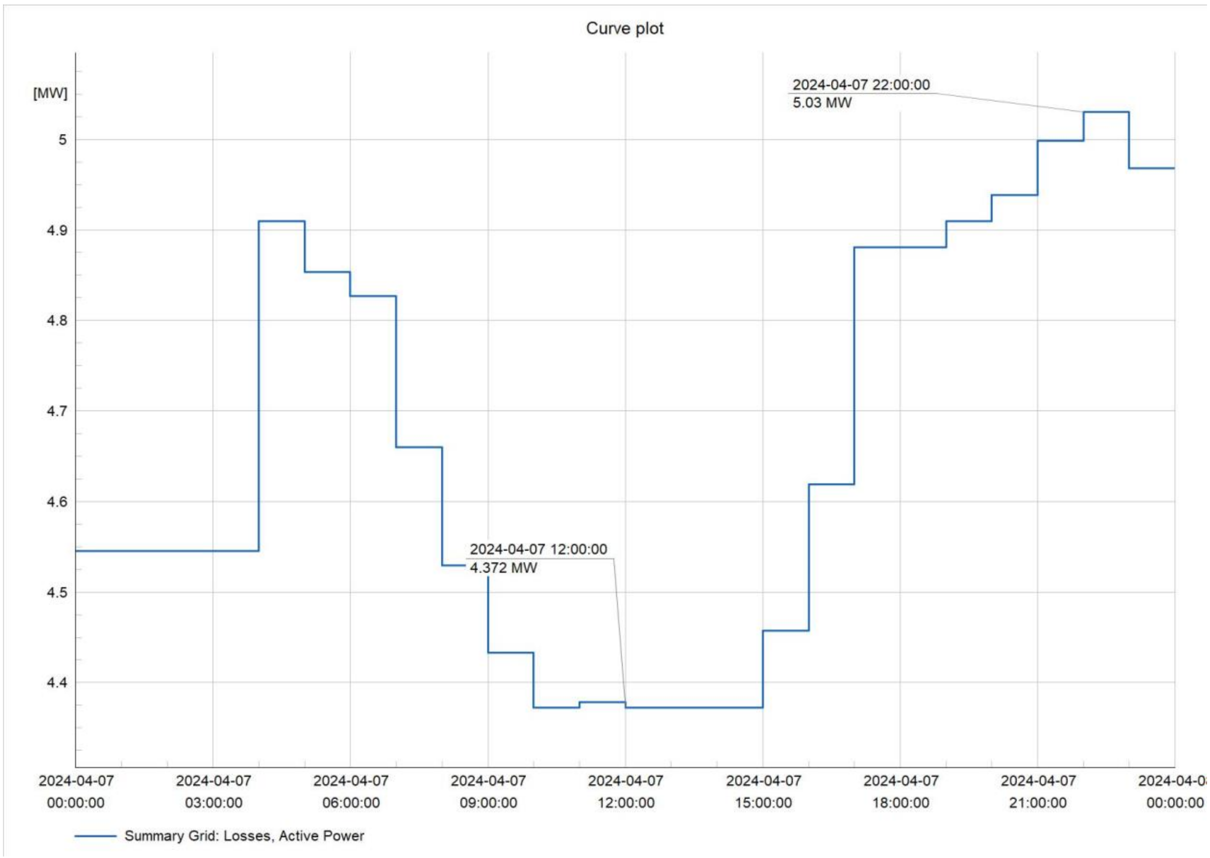



Curve plot
[MW]
6.4
6.2
6
5.8
5.6
5.4
5.2
5
4.8
2024-04-07 22:00:00
6.294 MW
2024-04-07 10:00:00
4.756 MW
2024-04-07 00:00:00
2024-04-07 03:00:00
2024-04-07 06:00:00
2024-04-07 09:00:00
2024-04-07 12:00:00
2024-04-07 15:00:00
2024-04-07 18:00:00
2024-04-07 21:00:00
2024-04-0 00:00:00
Summary Grid: Losses, Active Power

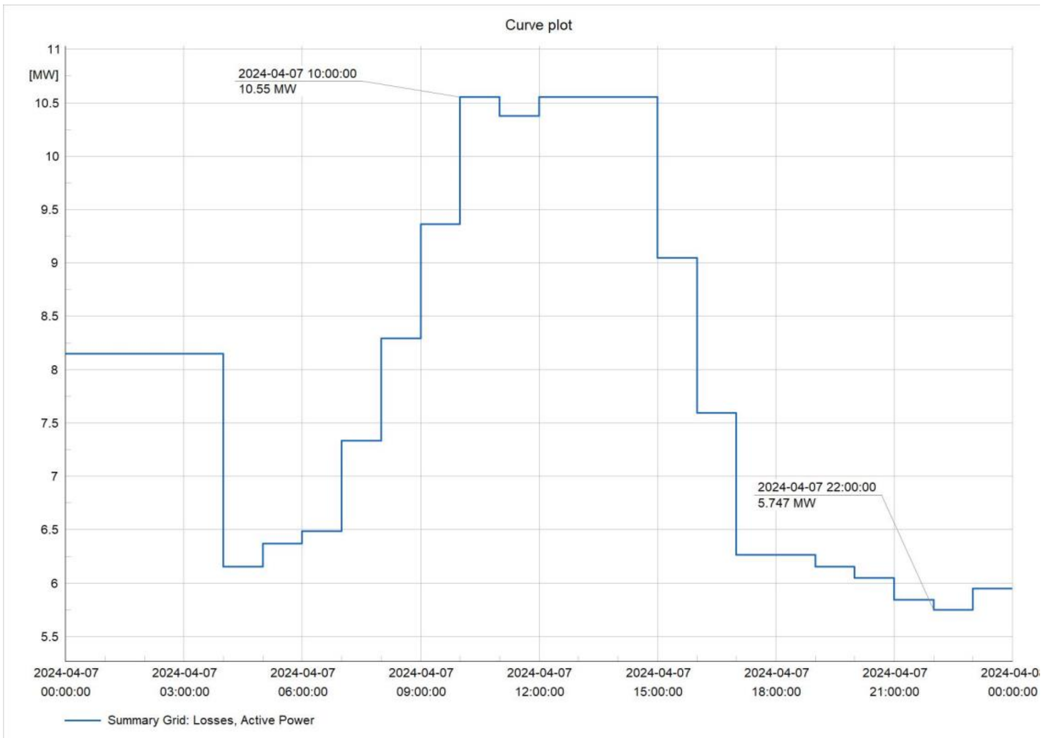
Case 2:

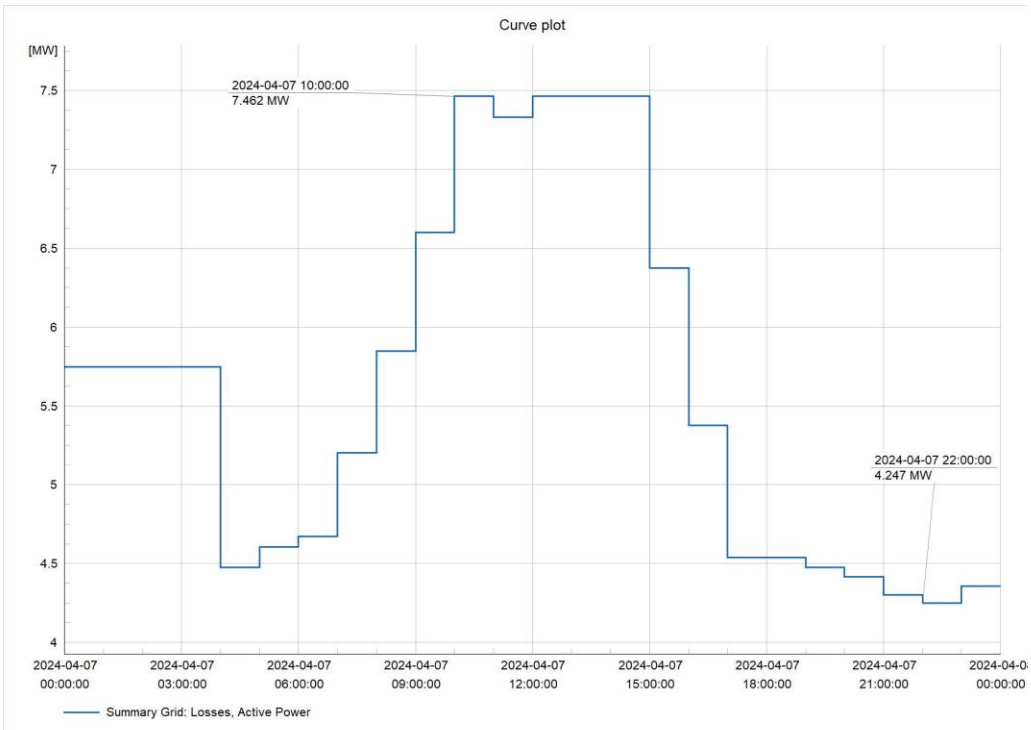

Curve plot
[MW]
4.9
4.8
4.7
4.6
4.5
4.4
4.3
2024-04-07 10:00:00
4.907 MW
2024-04-07 22:00:00
4.355 MW
2024-04-07 00:00:00
2024-04-07 03:00:00
2024-04-07 06:00:00
2024-04-07 09:00:00
2024-04-07 12:00:00
2024-04-07 15:00:00
2024-04-07 18:00:00
2024-04-07 21:00:00
2024-04-0 00:00:00
Summary Grid: Losses, Active Power

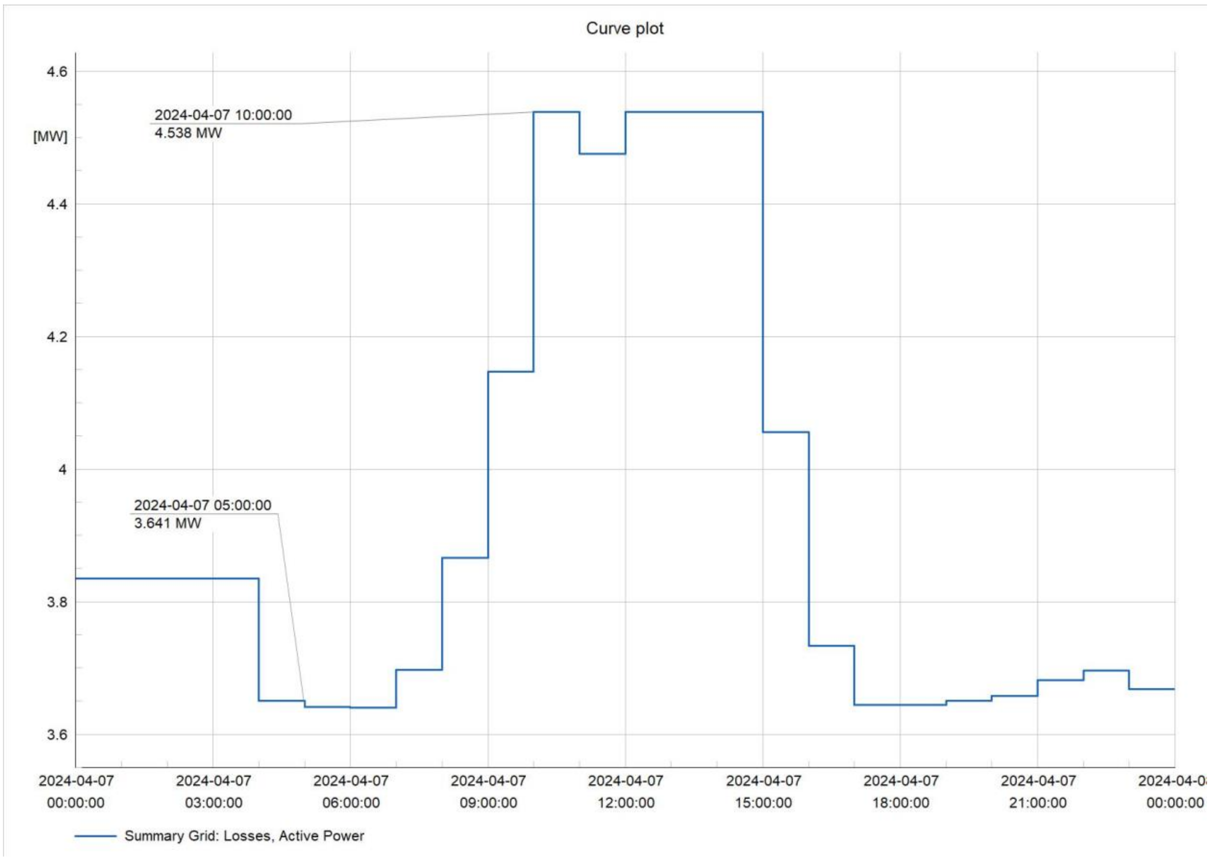

Curve plot
[MW]
4.5
4.48
4.46
4.44
4.42
4.4
4.38
4.36
4.34
2024-04-07 10:00:00
4.49 MW
2024-04-07 07:00:00
4.343 MW
2024-04-07 00:00:00
2024-04-07 03:00:00
2024-04-07 06:00:00
2024-04-07 09:00:00
2024-04-07 12:00:00
2024-04-07 15:00:00
2024-04-07 18:00:00
2024-04-07 21:00:00
2024-04-0 00:00:00
Summary Grid: Losses, Active Power


Curve plot
[MW]
5
4.9
4.8
4.7
4.6
4.5
4.4
2024-04-07 22:00:00
5.03 MW
2024-04-07 12:00:00
4.372 MW
2024-04-07 00:00:00
2024-04-07 03:00:00
2024-04-07 06:00:00
2024-04-07 09:00:00
2024-04-07 12:00:00
2024-04-07 15:00:00
2024-04-07 18:00:00
2024-04-07 21:00:00
2024-04-0 00:00:00
Summary Grid: Losses, Active Power

Case 3:

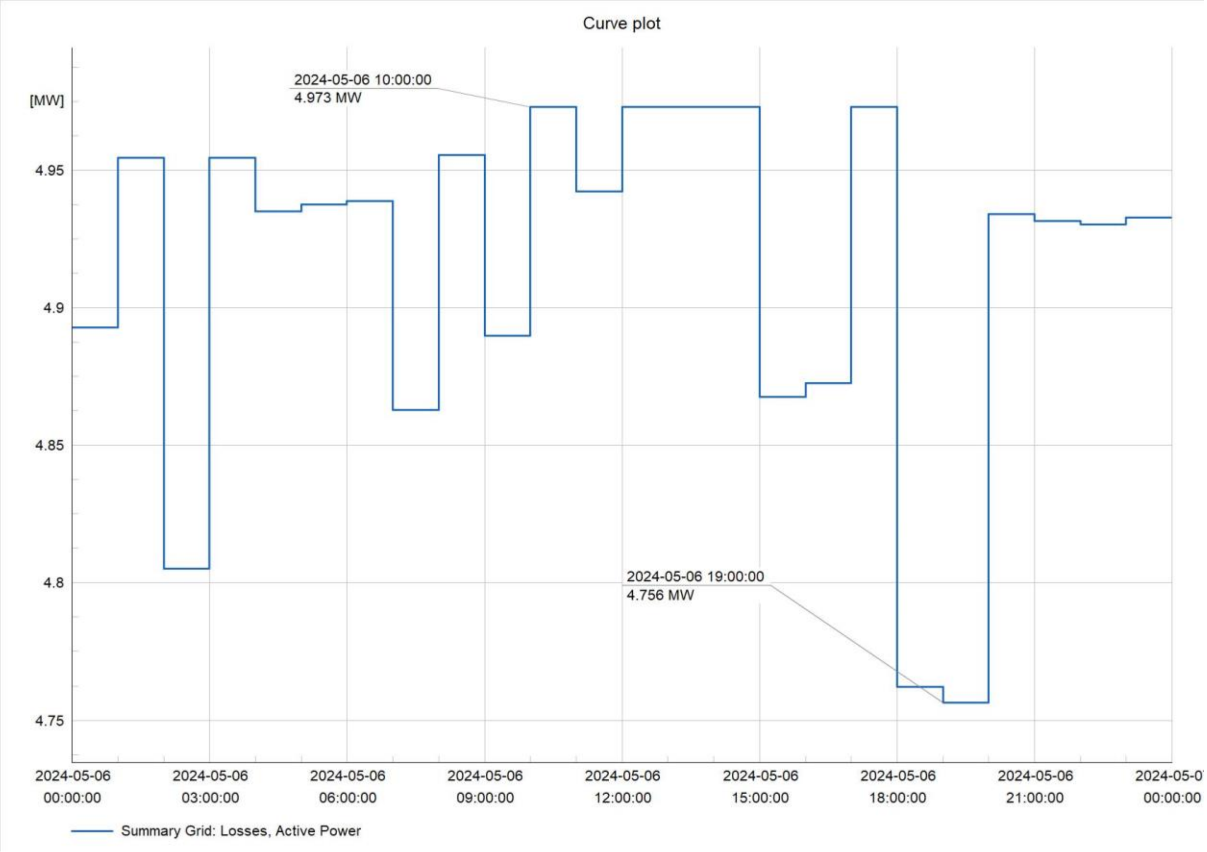

Curve plot
[MW]
11
10.5
10
9.5
9
8.5
8
7.5
7
6.5
6
5.5
2024-04-07 10:00:00
10.55 MW
2024-04-07 22:00:00
5.747 MW
2024-04-07 00:00:00
2024-04-07 03:00:00
2024-04-07 06:00:00
2024-04-07 09:00:00
2024-04-07 12:00:00
2024-04-07 15:00:00
2024-04-07 18:00:00
2024-04-07 21:00:00
2024-04-0 00:00:00
Summary Grid: Losses, Active Power

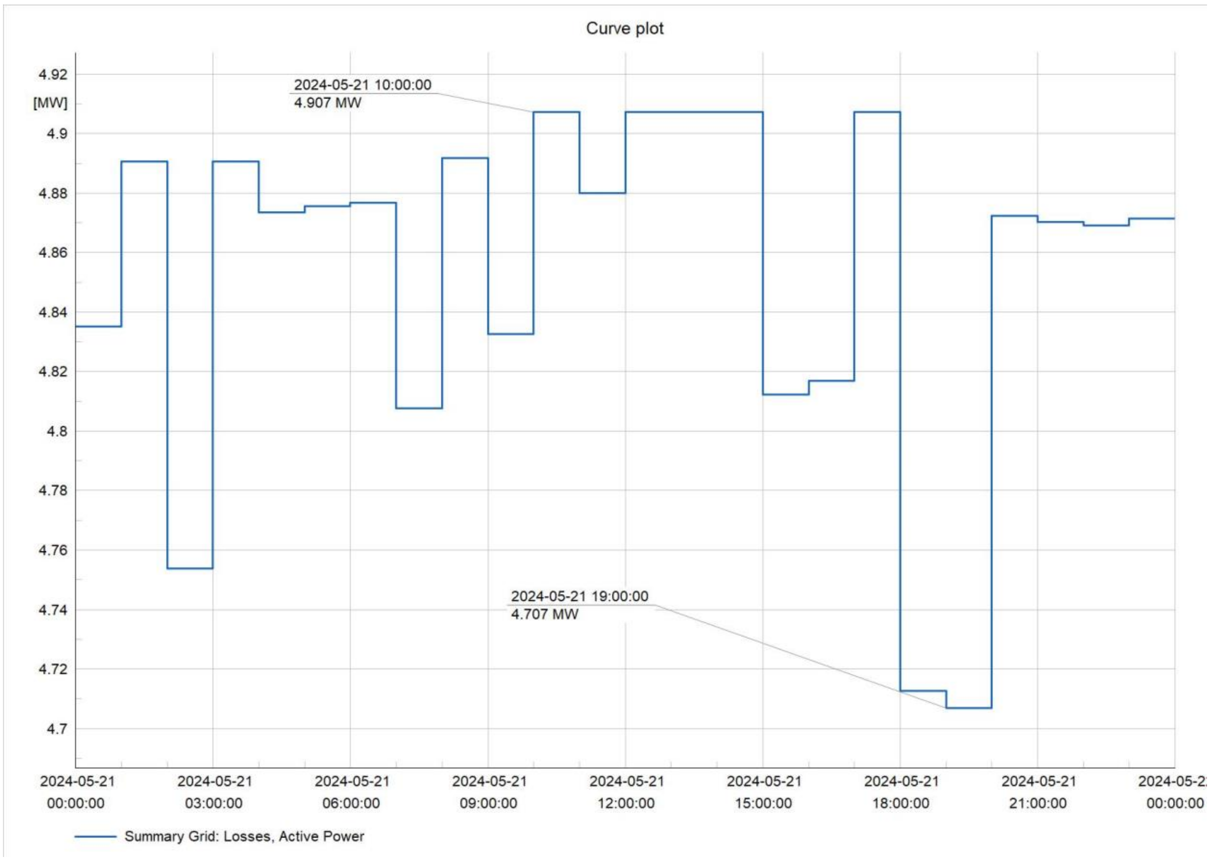

Curve plot
[MW]
7.5
7
6.5
6
5.5
5
4.5
4
2024-04-07 10:00:00
7.462 MW
2024-04-07 22:00:00
4.247 MW
2024-04-07 00:00:00
2024-04-07 03:00:00
2024-04-07 06:00:00
2024-04-07 09:00:00
2024-04-07 12:00:00
2024-04-07 15:00:00
2024-04-07 18:00:00
2024-04-07 21:00:00
2024-04-0 00:00:00
Summary Grid: Losses, Active Power

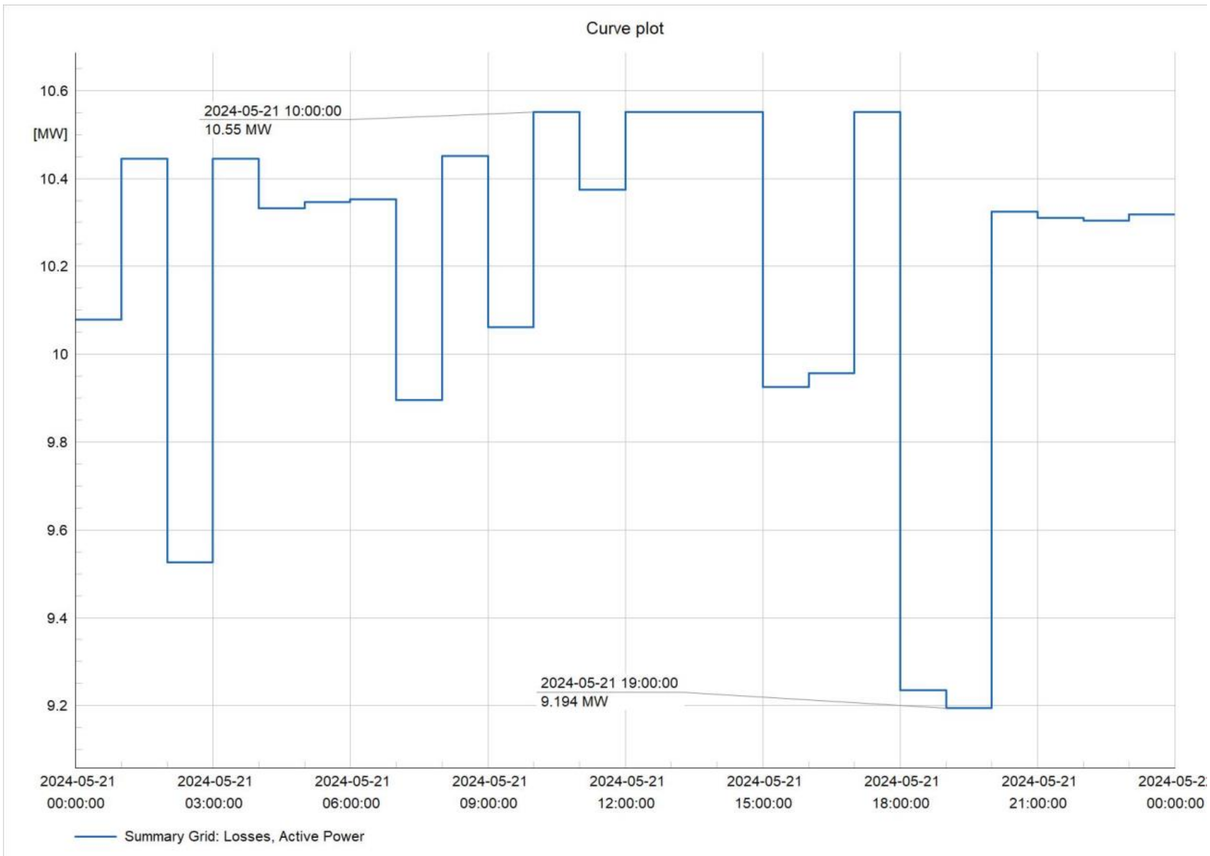

Curve plot
[MW]
4.6
4.4
4.2
4
3.8
3.6
2024-04-07 10:00:00
4.538 MW
2024-04-07 05:00:00
3.641 MW
2024-04-07 00:00:00
2024-04-07 03:00:00
2024-04-07 06:00:00
2024-04-07 09:00:00
2024-04-07 12:00:00
2024-04-07 15:00:00
2024-04-07 18:00:00
2024-04-07 21:00:00
2024-04-0 00:00:00
Summary Grid: Losses, Active Power

Appendix C

The simulation results for the load curve with prioritized workloads are presented below:

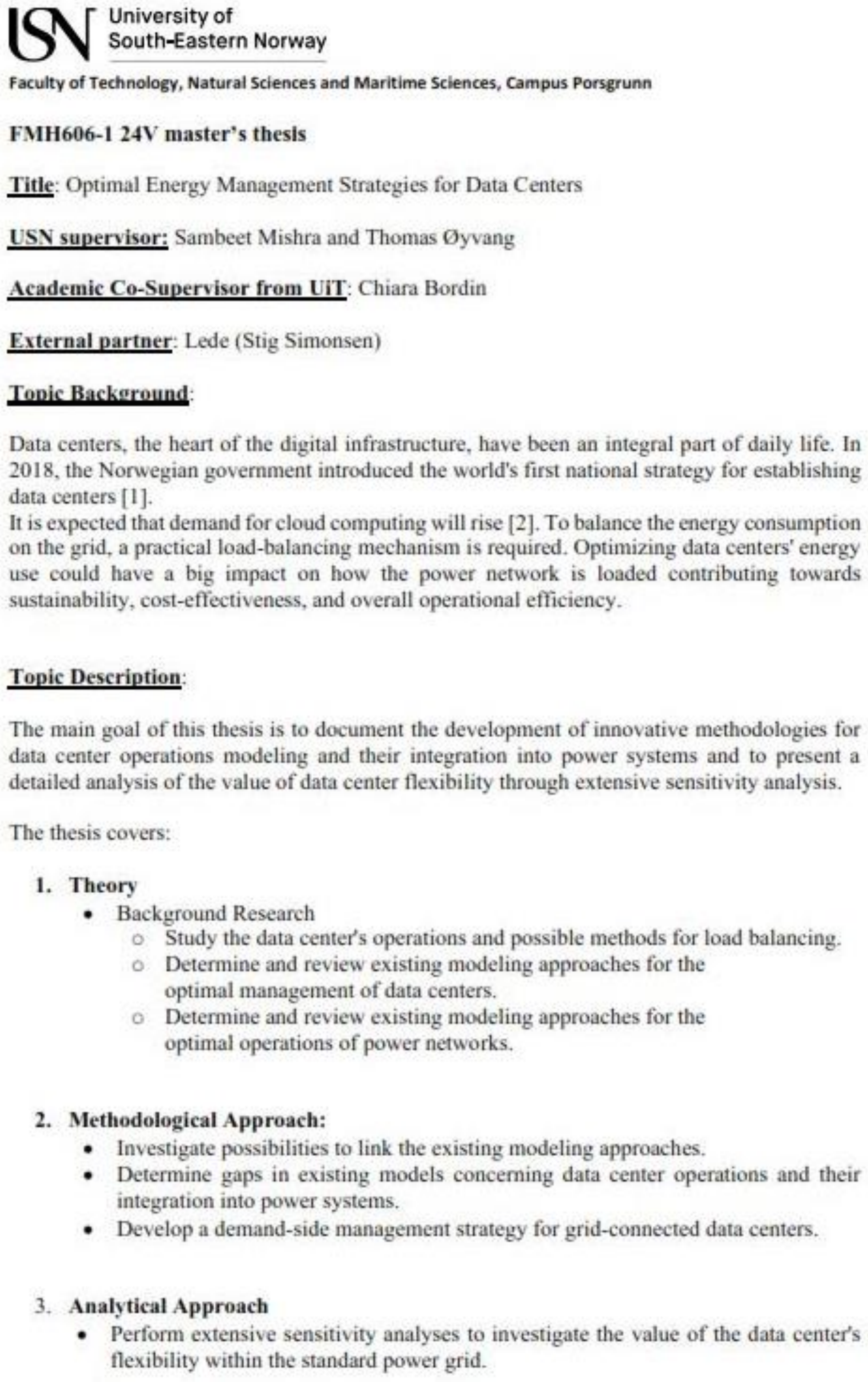





Curve plot
[MW]
10.6
10.4
10.2
10
9.8
9.6
9.4
9.2
2024-05-21 10:00:00
10.55 MW
2024-05-21 19:00:00
9.194 MW
2024-05-21 00:00:00
2024-05-21 03:00:00
2024-05-21 06:00:00
2024-05-21 09:00:00
2024-05-21 12:00:00
2024-05-21 15:00:00
2024-05-21 18:00:00
2024-05-21 21:00:00
2024-05-2 00:00:00
Summary Grid: Losses, Active Power

Appendix F

Project Thesis Description



University of South-Eastern Norway

Faculty of Technology, Natural Sciences and Maritime Sciences, Campus Porsgrunn

FMH606-1 24V master's thesis

Title: Optimal Energy Management Strategies for Data Centers

USN supervisor: Sambeet Mishra and Thomas Øyvang

Academic Co-Supervisor from UiT: Chiara Bordin

External partner: Lede (Stig Simonsen)

Topic Background:

Data centers, the heart of the digital infrastructure, have been an integral part of daily life. In 2018, the Norwegian government introduced the world's first national strategy for establishing data centers [1].
It is expected that demand for cloud computing will rise [2]. To balance the energy consumption on the grid, a practical load-balancing mechanism is required. Optimizing data centers' energy use could have a big impact on how the power network is loaded contributing towards sustainability, cost-effectiveness, and overall operational efficiency.

Topic Description:

The main goal of this thesis is to document the development of innovative methodologies for data center operations modeling and their integration into power systems and to present a detailed analysis of the value of data center flexibility through extensive sensitivity analysis.

The thesis covers:

1. **Theory**
   - Background Research
     - Study the data center's operations and possible methods for load balancing.
     - Determine and review existing modeling approaches for the optimal management of data centers.
     - Determine and review existing modeling approaches for the optimal operations of power networks.

2. **Methodological Approach:**
   - Investigate possibilities to link the existing modeling approaches.
   - Determine gaps in existing models concerning data center operations and their integration into power systems.
   - Develop a demand-side management strategy for grid-connected data centers.

3. **Analytical Approach**
   - Perform extensive sensitivity analyses to investigate the value of the data center's flexibility within the standard power grid.

Signatures:

Supervisors:

S Mishra.

Sambeet Mishra

Student:

SANDHYA BOHARA

References:

[1] M. E. a. D. Consulting, «how-to-establish-a-data-center-in-norway,» *how-to-establish-a-data-center-in-norway,* p. 22.

[2] M. K. &. C. Company, «Investing in the rising data center economy,» [Internett]. Available: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy. [Funnet 18 January 2024].