MSc Program
Engineering Management

ACADEMY FOR CONTINUING EDUCATION

Cooling energy saving mechanisms for Data Centers.

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by
Em.O.Univ.Prof. Dipl.-Ing. Dr.techn. Dr.h.c.mult. Kopacek Peter

Ahadulla Nazurmetov, BSc

12038050

Vienna, 10.03.2022

Affidavit

I, **AHADULLA NAZURMETOV, BSC**, hereby declare

1. that I am the sole author of the present Master's Thesis, "COOLING ENERGY SAVING MECHANISMS FOR DATA CENTERS.", 79 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and
2. that I have not prior to this date submitted the topic of this Master's Thesis or parts of it in any form for assessment as an examination paper, either in Austria or abroad.

Vienna, 10.03.2022

____

Signature

Acknowledgement

I would like to express my deep and sincere gratitude to all those who have assisted me in putting these concepts, which are much above the level of simplicity, into something solid.

I'd want to convey my heartfelt appreciation to Professor Peter Kopacek for not only overseeing my thesis and providing helpful criticism and suggestions but also for directing the full master's program. It was a great privilege and honour to work and study under his guidance.

Additionally, I'd want to express my gratitude to my colleagues and classmates, whom I've grown to know rather well during the program. I've liked learning, participating in spirited debates, and sharing a drink with all of you after class.

Finally, but certainly not least, I am indebted to my wife, Günel Guliyeva, for being in my life, for her unconditional love, for her patience and tolerance during the duration of my master thesis.

Abstract

Globally, data centers' energy requirements are expanding at a fast pace, as are their environmental and economic impacts. Typically, data centers are quite an energy hungry. The annual cost of powering a big data center, including computers and cooling equipment, may be approximately $4 million. Energy prices may account for up to 60% of operational costs. According to recent research from the Department of Energy's Lawrence Berkeley National Laboratory, data centers in the United States use a significant amount of energy – around 70 billion kilowatt-hours each year. The electrical energy consumption for data centers in Germany is 16 billion kilowatt-hours per year – up 8% over the previous year in 2020. This increase is mostly attributable to the rapid rise of cloud computing capability. The energy used directly by servers is just one aspect of the picture. Almost every unit of energy used by a server is converted to heat in the building in which it is located, requiring air-conditioning systems to work harder than they would otherwise. Because IT systems often emit large amounts of heat and have a high need for temperature and humidity management in data centers, a cooling and air conditioning system is required. According to a recent study, refrigeration, and air conditioning systems account for around 40% of overall data center power use. As a result, decreasing the energy consumption of the air conditioning system is critical for overall data center energy efficiency.

This thesis describes in detail the cooling systems used in the data center's industry, as well as the modern immersion cooling system. A comparative analysis of the effectiveness of immersion cooling in comparison with the traditional air-cooled system is conducted. On the example of a real data center, metrics analyzes in detail, showing how effectively the data center is optimized. The last chapter provides in detail the waste heat recuperation technology.

This thesis provides practical information to the data center engineers regarding the cooling system design and supports them to achieve better cooling energy efficiency.

Table of Contents

1 INTRODUCTION

A data center is a facility with a physical space dedicated to the storage of networking and computing infrastructure to enable shared access to applications and data. Similarly to how programs work on laptops, in order to access anything on the internet requires the use of a computer. These machines are collectively referred to as servers. They are similar to laptops but lack a screen and keyboard and must be situated near an internet connection, electricity, and cooling. These facilities are referred to as data centers. These facilities might be as tiny as 30 m2 cabinets or as large as 120.000 m2 "hyper-scale" warehouses.

When you use an internet service, you are connecting to one of the many millions of servers housed in one of the thousands of data centers spread across the world. Data centers have the aforementioned computer systems and the personnel responsible for their maintenance. The data centers requirement includes power systems with options of backup power, cooling systems, fire safety systems, security systems. Data centers come in a variety of sizes, from a small single rack capacity to clusters of geographically allocated buildings. They are all critical business assets where businesses frequently invest in and deploy cutting-edge data center technologies.

The history of the data center evolution shows that the data centers have grown organically, beginning as IT departments or research centers. It is rather difficult to ascertain a particular date in history when a data center was created. It may be gleaned from the papers pertaining to the establishment of the data center, from when the first server was deployed or from when it started operations.

Fig. 1 shows the typical overview of the Data Center (Datacenter at CERN, 2016)

1.1 Electrical system

IT equipment and its associated equipment are extremely energy-intensive. Electricity expenses are a significant operational expenditure of data center facilities and usually represent more than 10% of the overall cost of data center operation. Also, it is critical to have reliable electricity for IT equipment as well as other data center equipment such as cooling or security systems. One or more uninterruptible power supply and diesel generators comprise a backup power system. In order to avoid single points of failure, electrical systems are often fully duplicated, and a typical data center is usually fed from a dual power source. There are multiple redundancy configurations that exist for the data centers that are frequently referred to as N+1 or 2N redundancies. The special kind of change over system that is named a static switch is used to provide quick switchover between two different power supplies in the case of a power outage. The term "power use effectiveness" is used to compare the electro-efficiency of different data centers. Along with servers, a cooling system is big energy-consuming equipment. Less than 33% of the energy consumption only falls under IT equipment. The following are the most often used components of electrical systems:

- Uninterruptible power supplies
- Building and power management systems
- Backup diesel generators
- Medium and low voltage switchgear
- power distribution systems
- automatic transfer and static switches

1.2 Medium voltage system

Medium voltage lines are used in hyper-scale data centers to distribute power across different step-down transformers. This results in less complexity of the data center infrastructure. The complexity is one biggest issues that are negatively affecting to the resiliency of the data centers. As an example of 1 MW power, if to distribute it over 4 kV system, it will require 9 times fewer cables and busways to compare with 400V system. This is a significant reduction of the infrastructure’s complexity.

To ensure continued operation in the event of a widespread power loss, data centers are often equipped with at least one diesel or gas backup generator. Both the local utility provider and the backup generator supply energy at a medium voltage.

1.3 Low voltage systems

As is mentioned above, the medium voltage is converted to the low voltage in the step-down transformers. The low voltage is then distributed over the cables and busways to the main distribution boards and to uninterruptable power supplies which are, in turn, provide emergency power in the event that the input power source fails and protects key components from voltage spikes, harmonic distortion, and other power issues.

During a power outage, the majority of UPS systems are intended to deliver power for at least five minutes at maximum load. The backup generators have ample time during those five minutes to start and take over the load from the UPS system.

1.4 Cables routing

Nowadays data centers communicate mostly via networks that support Internet protocols and specialized protocols for interconnecting IT equipment. Data centers are equipped with a network of switches that provide communication between the computer equipment in the data center and the rest of the world. Often, Internet connection redundancy is achieved by utilizing two or more upstream Internet service providers. Additionally, network security features like firewalls, VPN gateways, and intrusion detection systems are typically used. Usually, all the network cables are installed over the IT equipment, thorough the cable trays.

1.5 Power management systems

The energy management system is intended to monitor, measure and manage the electrical power consumption over the data center. The system generates reports on efficiency and failures to the millisecond, and the program may forecast prospective problems based on data patterns, allowing the user to avoid them. The energy management system is a specialized supervisory control and data acquisition system designed specifically for electricity distribution. Numerous services done by this system are comparable to those performed by a SCADA system, including the following:

- Individualized graphics

- Displays in real-time
- Alarming
- Trending and reporting

1.6 Fire safety system

Data centers provide a particular type of fire hazard. Data centers, which are densely packed with electrical components, generators, wires, connections, and backup power sources, provide a variety of potential fire threats. Given the sensitivity of the majority of the equipment in a data center, these facilities require customized specific hazard solutions. When researching data center fire protection standards, it's critical to understand the requirements that your data center must satisfy and the fire prevention and suppression systems that can support them. Due to the multiple fire threats, fire protection is required at three distinct levels — building, room, and rack.

The first level – building- the fire protection system at this level is aimed at ensuring the safety of the building and its occupants. Several methods are employed in data centers to achieve building-level fire protection:

- Sprinkler system
- Manual fire extinguishers
- Passive design elements: fire rated elements and firewalls that act passively to retard the spread of fire which allow for the temporary confinement of a fire to a portion of the facility in the event that the active fire protection systems fail, such as when the door is left open or when they are not installed. These firewalls, however, are frequently insufficient to safeguard heat-sensitive electronic equipment in critical facilities, as traditional firewall architecture is certified for flame penetration duration rather than heat penetration. Additionally, there are vulnerabilities in the server room's security of vulnerable entrance points, such as cable penetrations, chilled water pipelines and air exchange ducts.

At the room level protection, which is the second level of fire protection, the fire suppression system is used. There are different types of fire suppression systems available on the market, here are the main ones:

- Wet or water mist sprinkler system. Due to the fact that the water is just as detrimental to the equipment as a fire would be, this type of system is not very common in data center operations. Additionally, if a wet system were to leak or drip, the resulting damage would be severe.
- Pre-action system. This system also utilizes the water to extinguish a fire, but two distinct events must occur in order for the system to operate. The detecting system must first detect the presence of a growing fire and then open the pre-action valve. This enables for the flow of water into the system's plumbing, thereby creating a wet pipe sprinkler system. Second, each sprinkler head must be released to allow water to pour onto the flames.
- Clean agent system commonly used to extinguish and control a fire in data centers critical spaces. A clean agent is the form of an agent that is not electrically conducting, liquid, or gaseous, and that evaporates without leaving a residue. Fire suppression systems also consist of inert gas that is held in a container and released in response to the detection of a fire.

Apart from the fire suppression system, the fire safety at the room level also comprises fire detection systems. VESDA (very early smoke detection apparatus) is the most commonly used detection system in the data center industry that is able to detect the smoke of a potential fire at very early stages.

The rack-level protection is used to detect and extinguish the fire at the early stages. Usually, the automatic fire suppression system is installed within the rack.

1.7 Physical Security of a data center

Physical risks to data centers' components must be mitigated. A safe location, physical access restrictions on the building, and monitoring systems all contribute to the physical security of a data center facility. While it is more common to attack a server or network remotely in order to steal and access data nowadays, physical assaults may be just as hazardous. Because firms that deal with sensitive and highly secret data are required to have cutting-edge online security measures, criminals frequently attempt to physically access the data center's weak areas in order to steal the data. Physical access to the location is often restricted to authorized employees through bollards and mantraps. Almost often, video monitoring and permanent security officers are present.

1.8 Data centers standards

As a result of the foregoing, we may deduce that data centers require wires and cables for power supply and fiber-optic communications, cooling systems (heating, ventilation, and air conditioning), temperature conditions, and fire safety, as well as physical security systems. All of these systems and technologies were accomplished through the development of technical solutions. At the moment, standards have been established for the equipment of data centers. For the proper placement and installation of expensive and complex IT equipment in data centers, the hyper-scale companies rely on a variety of ongoing research and development, and standards in the fields of telecommunications, power consumption, cooling systems, fire prevention, and suppression, and physical security systems, among others. Industrial solutions required for data centers are applied in a wide variety of different economic sectors. For example, ASHRAE was created in 1959 via the merging of The American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) founded in 1894 and The American Society of Refrigerating Engineers (ASRE) established in 1904. (ASHRAE Journal, 1972) These groups' suggestions and standards are taken into account when developing data centers, with an emphasis on building management systems, power usage efficiency, air quality, and industrial reliability and sustainability.

The administration of associations, research organizations, and individual enterprises facilitates the development of technologies, for example:

- Association for Computer Operations Management that was established in 1980
- The Open Data cent Alliance, established in 2010
- EUDCA (European Data Centre Association), was formed in 2011

DIN EN 50600 is the first cross-border standard in Europe, which uses a holistic approach to provide comprehensive specifications for the planning, construction, and operation of a data center. It defines requirements for planning the trades of building construction, electrical supply, air conditioning, cabling, security systems and sets criteria for the operation of data centers. DIN EN 50600, created by the European standardization company CENELEC (European Committee for Electrotechnical Standardization), offers various degrees of freedom and, up to a certain point, sees itself as a modular system. First and foremost, DIN EN 50600 is a standard that is used in the construction of new data centers. It defines the need for expert opinions and analyzes in the run-up to the planning and construction work. While the ISO management standards are also relevant in the data center environment, e.g. B. ISO/IEC 27001, have a focus on the organizational and procedural level, the requirements of DIN EN 50600 focus on physical security. The standard, which is valid throughout Europe, is something special among a large number of guidelines and best practices in that the results were created in a Europe-wide standardization and coordination process.

The Tier level categorization system is the most critical component of data center standardization. The Data Center Tier Performance Standards are a collection of user-defined standards that outline the requirements for the design and administration of the data center in order to achieve a certain degree of availability. The Tier Level Classification system defines the industry-wide standards and requirements for performance management of data center critical infrastructure topology and operational plans which many data center industries consultants, and design experts base their approach to rating data center projects by comparing design against performance. (UptimeInstitute, 2021)

The Uptime Institute, founded in 1993, is an independent, non-profit data center research, teaching, and consulting organization dedicated to enhancing data center performance and efficiency via cooperation and innovation. For almost three decades, data center owners and operators have used the Uptime Institute's Tier Standard and Tier Classification System. The Uptime Institute services the data center industry's whole ecosystem of stakeholders. This collaborative approach, along with the Uptime Institute's capacity to identify global trends and communicate directly with owners, leads in solutions and invocations that are not constrained by geographical boundaries, benefiting the global data center business. Since its inception in 1993, the Institute has started the development of knowledge communities with the goal of increasing the dependability and uninterruptible availability of data center facilities and IT companies. The activities of Uptime Institute began with UUUG (Uninterruptible Uptime Users Group) which is in 1989 continued to fill a void in the data center community for the user's capacity to freely communicate knowledge through the open forum structure of the group meetings. Members have steadily increased site uptime through a collaborative learning experience. Selection of site and evaluation of performance are critical parts of data center performance evaluation.

Location	Earthquake zone Weather disasters The proximity of Hazardous areas Airports, highways, railways proximity
Infrastructure	Electricity availability Outages Utilities The capability of future expansion Power feeders
Economics	Land Construction Labour Cost-effectiveness Communications
Communications	Availability of Diverse Carriers Services Physical Security Alarms and Monitoring

Table 1 shows the main site selection criteria (Balodis & Opmane, 2017)

Electrical	Power Backbone – Redundancies Utility Service UPS Systems UPS Batteries Power Generator Critical Power Load Bank Distribution Protection Lightning Grounding

Mechanical	Cooling system Transformers and UPS cooling Mechanical system
Support services	Fire detection Fire suppression Physical security Alarms and Monitoring

Table 2 shows the main site performance evaluation (Balodis & Opmane, 2017)

The Tier I data center is comprised of non-redundant electrical and mechanical systems and a single non-redundant distribution line (for electrical and cooling supply) that serves the site's IT equipment. The data center is equipped with cooling and power distribution but may or may not include a backup power supply or power generator. For routine maintenance or emergency repairs, the crew must completely shut down operations. A data center interruption will occur as a result of operational faults or spontaneous breakdowns of infrastructure components. Clients of tier I data centers might anticipate up to 28.8 hours of downtime per year due to a lack of redundant systems. A Tier I data center, for example, may be appropriate for small firms where IT is used to support internal business activities.

Tier II data center infrastructure has all of the characteristics of tier I infrastructure, but with the addition of backup capabilities and has redundant electrical and mechanical systems but just one non-redundant distribution busway for electrical system supplying the site's IT equipment. They include uninterruptible power supplies (UPS) and power generators, with Need plus One (N+1) redundancy, but with a single power route. Maintenance of the site's essential electrical bus ducts and other infrastructure components will necessitate the shutdown of IT equipment operations. As a result of the higher uptime compared to a lower-rated data center, tier II customers might anticipate up to 22 hours of downtime per year. Tier II data center may be suitable for internet-based businesses without imposing significant financial penalties for meeting quality of service promises.

The Tier III data center is equipped with multiple independent electrical and mechanical distribution routes and systems serving the site's IT equipment. During preventative maintenance or equipment replacement, a tier III plant does not need a complete shutdown. Despite that, the data center is composed of multiple electrical and mechanical distribution routes, but only one route is redundant and maintainable. Only one distribution route is available at any one moment to service the IT equipment. This design enables any scheduled site infrastructure maintenance and breakdown correction activities to occur without interfering with the operation of the IT equipment. Tier II configurations show a noticeable increase in availability when compared to lesser classifications. Clients of a Tier III data center might expect annual downtime of up to 1.6 hours. Tier III applications include businesses that operate in different time zones or depend on information technology to support the business.

The Tier IV data center with redundant electrical and mechanical systems and numerous distribution routes supporting the site's IT equipment, including networking, cooling, diesel generators, fuel systems. Every component is fault-tolerant, dual-powered, and appropriately placed to ensure compatibility with the topology of the site's architecture. A Tier 4 data configuration can assure that customers are not down for more than 0.8 hours each year. Tier IV criteria apply to businesses that need exceptionally high availability for continuing.

The total Tier Level is often determined by the lowest Tier rating or weakest component. For instance, a data center may have a Tier III rating for the electrical system but a Tier II rating for a mechanical system, resulting in the data center's total Tier rating of II. In actuality, a data center's infrastructure may have varying tier ratings.

Configuration	Tier I	Tier II	Tier III	Tier IV
Availability	99.67%	99.75%	99.98%	99.99%
Unplanned outages	28.8h per year	22h per year	1.6h per year	0.8h per year
Redundancy	No	Partial N+1	N+1	2N or higher
Concurrently maintainable	No	No	Yes	Yes

Table 3 shows the Datacenter Tiers compared (Uptime Institute, 2009)

1.9 Data center development history

The term "data center" refers to the use of information technology in the industry. The early stages of data center development are inextricably linked to mainframe computer maintenance. The first unique computer systems were developed and disposed of in research labs. The earliest phase of data centers is associated with such computer disposal and maintenance that occurred outside of research labs in at least a few cases. The earliest industrially built computers and data centers drew on decades of mechanical data processing expertise with various tabulators, adding machines, and other devices. The first computers were room-sized equipment that took up a lot of space. Due to the difficulty of managing and maintaining this equipment, they were also segregated in specialized rooms. Historically, the commercial value of data center physical infrastructure was determined by two primary criteria: uptime and initial cost.

Engineers from the University of Pennsylvania created the first ENIAC data center in secret. In 1943, J. Presper Eckert and John Mauchly promised to overcome the speed limitations of electromechanical computing with their proposal to build a tube computer. Calculating more than a thousand times faster than Aiken's electromechanical ASCC was only possible with tubes. (IBM's ASCC introduction, 2022)

The ENIAC (Electronic Numerical Integrator and Computer) required 22 tubes to electronically provide one digit of a ten-digit decimal number (Computer History Museum, 2022). 220 tubes for a decimal number, 8800 tubes for 20 accumulators and again over 9,000 tubes for the multiplication, division, and input/output units. A total of 18,000 tubes had to be kept operational for the duration of a computing task. It was developed by the United States army to calculate artillery fire during World War II and was also utilized by scientists to construct the first thermonuclear bomb, which was finally launched on Hiroshima and Nagasaki in 1945.

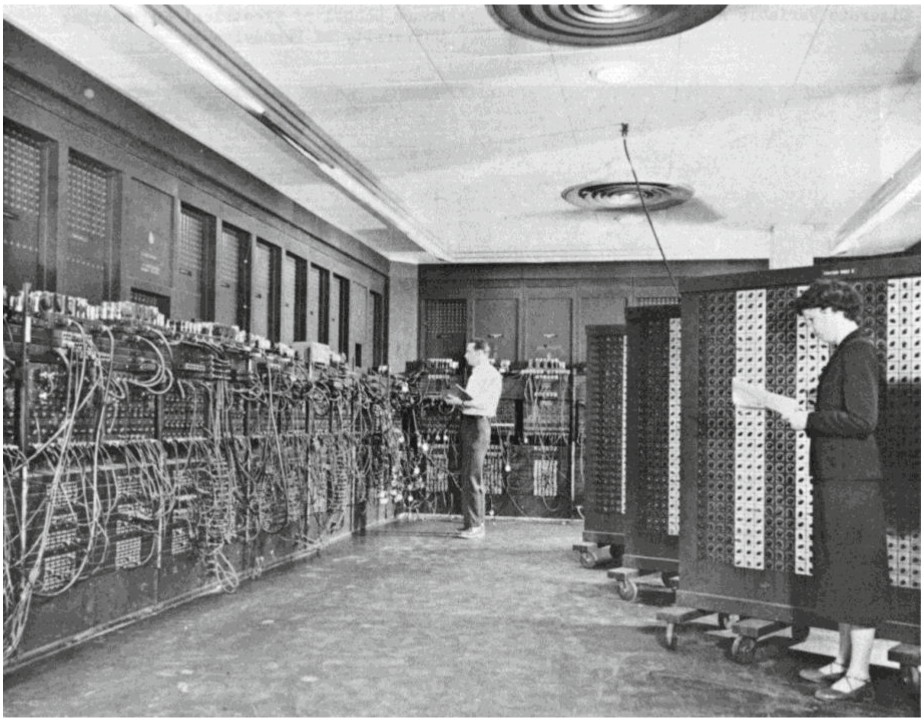


Fig. 2 shows the first system operating on ENIAC computer. "Historical Monograph: Electronic Computers Within the Ordnance Corps" The ENIAC (U. S. Army Photo, 2022)

In 1958, CERN erected its first computer, a massive vacuum-tube Ferranti Mercury. It ran at a millionth the speed of today's huge computers. While the Mercury took three months to build — and took up a large amount of space - its processing capability fell short of that of a current pocket calculator. Four magnetic drums each containing 32K 20 bits offered "mass" storage — insufficient to retain the data from a single proton-proton collision at the Large Hadron Collider. (CERN, 2022)

Throughout the 1960s and 1970s, data processing computers advanced even more intense. IBM introduced TRADIC, the world's first transistorized computer, in the early 1960s. This innovative paradigm facilitated the transition of data centers from the military to the commercial sector by obviating the requirement for expensive vacuum tube equipment. Additionally, TRADIC improved computing capabilities and features dramatically. Additionally, it reduced the size of computer systems, making them more easily integrated into multifunctional areas such as office buildings. NASA used these methods to land humans on the Moon.

Rapid invention and development in the 1960s and 1970s, along with a vast growth of computer businesses such as Intel, IBM, Sun Microsystems, and Xerox, foreshadowed the 1980s advent of personal computing. Worldwide, information technology began to become a significant economic contributor and has grown at an unparalleled rate.

UNIX established the norms for computer technologies and was a watershed moment in the development of data-driven technologies. It was based on a "client-server" architecture in which several computers were linked to a host server over the newly formed Internet. Users started interacting with servers housed in data centers across the globe over the Internet in the 1980s-1990s, creating the framework for the contemporary data center. The World Wide Web era saw a massive reorganization of global economies as a result of the development of the contemporary internet. As a result, it caused the global data center boom. The data center has developed into a crucial component of national security, internet infrastructure, and economic activity. Engineers from all around the globe have committed their efforts to perfecting the design of data centers.

Businesses need high-speed Internet access and constant operation on a never-before-seen scale. Because many small businesses could not afford to build such costly data centers infrastructure, international firms such as Amazon and Google started developing massive data centers that supply businesses with a variety of services and technical solutions. Massive proliferation continued throughout the 2000s-2010s. Thousands of new jobs were created as a result of these new data centers, which necessitated considerable modifications to local power infrastructures. While many were eager to seize the opportunity for increased earnings and enhanced local infrastructure, there has been much criticism around these new, hyperscale data centers. Hazards and harm to the environment swiftly followed.

Today, data centers remain a mystery to the majority of people. They protect the data of billions of users by being stored in safe places throughout the world. As with the secrecy surrounding the development of data centers during World War II, data centers remain heavily cloaked and concealed from the general public to this day.

1.10 Data centers energy demand

While the contemporary data center requires little human interaction, it consumes almost 3% of the world's energy production in aggregate. In 2020, the electricity demand of the data centers in Germany increased again significantly, from the below graph we can see that the data centers electricity usage increased by seven percent to 16 thousand GWh., despite the economic crisis caused by the corona pandemic. However, the effects of the coronavirus pandemic will affect different types of data centers in very different ways. Investments in smaller data centers operated by companies for their purposes have declined significantly. The general reluctance to invest in the crisis has a noticeable impact on such data centers. In contrast, many cloud providers have even benefited significantly from the corona pandemic in some cases. For example, the substantial increase in demand for video services, tools for online collaboration and online shopping are leading to significant market growth. (Bitkom e.V, 2021)

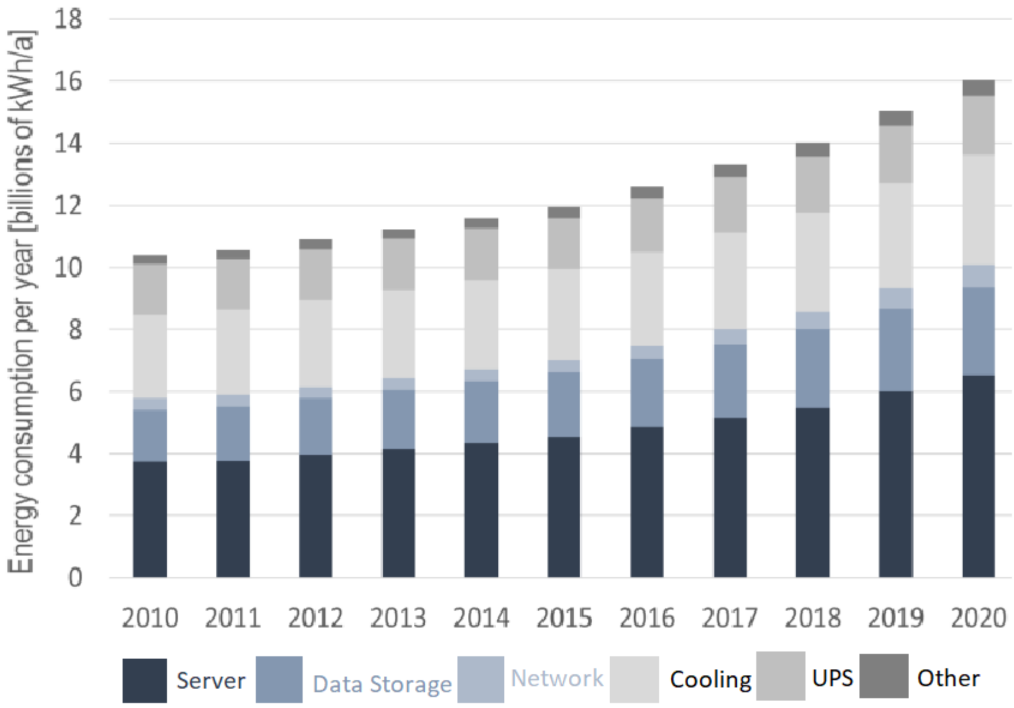


Fig. 4 shows the Data Centers Energy consumption (Borderstep Institut, 2020)

Edge data centers will also consume a growing amount of energy in the future. By 2025, Germany's edge data centers would likely consume 1.5 billion kWh of electrical energy. With the continued growth of 5G mobile phone networks and edge computing applications in sectors such as Industry 4.0, autonomous driving, and smart cities, edge data center energy consumption is estimated to reach around 4.5 billion kWh/a by 2030.

The expensive cost required for cooling infrastructure is one of the primary reasons enterprises shift away from on-premises data centers and toward colocation. When it comes to cooling IT equipment, the majority of private data centers and telecom closets are highly inefficient. Additionally, they lack the monitoring capabilities of colocation data centers, making it more difficult to properly optimize equipment for cooling demand reduction. Much power means a lot of heat.

There is no doubt that incorrectly calculated and designed data center cooling can lead to excessive heat; in turn, this causes severe strain on servers, storage devices, and networking equipment. This can result in downtime, damage to vital components, and decreased hardware lifespan, resulting in more significant capital costs. Inefficient cooling systems can drastically raise operational expenses.

1.11 Summary

In this chapter, in introductory part of our report, we examined in detail the main systems and components, the history and global development of data centers. Data center standards and concepts that govern how data centers are categorized were highlighted. The consumption of electricity was also considered, as well as forecasts for an increase in the demand for electricity, which by several times will exceed the current level. The Data centers energy consumption is only anticipated to grow as cloud computing, edge computing, the Internet of Things (IoT), artificial intelligence (AI), and other digital transformation technologies gain traction. Technological advancements will be canceled out by the ever-increasing amounts of computation and storage necessary to meet consumer and commercial needs. All this once again shows the urgent need for the development of technologies in the field of energy saving. It is a well-known fact that data centers are one of the largest consumers of electricity in the world. More and more countries are switching to alternative energy sources in connection with the expected global warming, which was noted in the Paris declaration on climate change, and therefore the topic of energy saving is acute for many hyperscale cloud services providers.

2 State of the Art

Data centers cooling systems

The main role of the data center’s cooling systems is to provide an appropriate environment that is required for the proper functioning of IT equipment. In order to do this, heat generated by IT equipment must be removed and transferred to a heat sink. The cooling system in the majority of data centers must work constantly and reliably. Operating a data center consumes a lot of energy, and frequently the cooling solutions consume as much as the computers they serve. However, with a well built and maintained cooling system, it is energy consumption may be reduced to a fraction of that of the IT equipment. Considering the critical nature of data center cooling infrastructure, it is worth delving into several existing and emerging data center cooling solutions.

2.1 Chilled water system

A chilled water system assists in cooling the entire data center or several rooms independently. In commercial settings, each room frequently has a single or many chilled water systems that function in conjunction with the air distribution system, such as Air Handling Units (AHU) or Computer Room Air Handler (CRAH). The chilled water system uses a mixture of glycol (35%) with water to remove the heat from data halls. The cooled water absorbs the heat and deposits it in the chiller as it flows through coils. After returning to the chiller, the water is combined with condenser water running via a cooling tower.

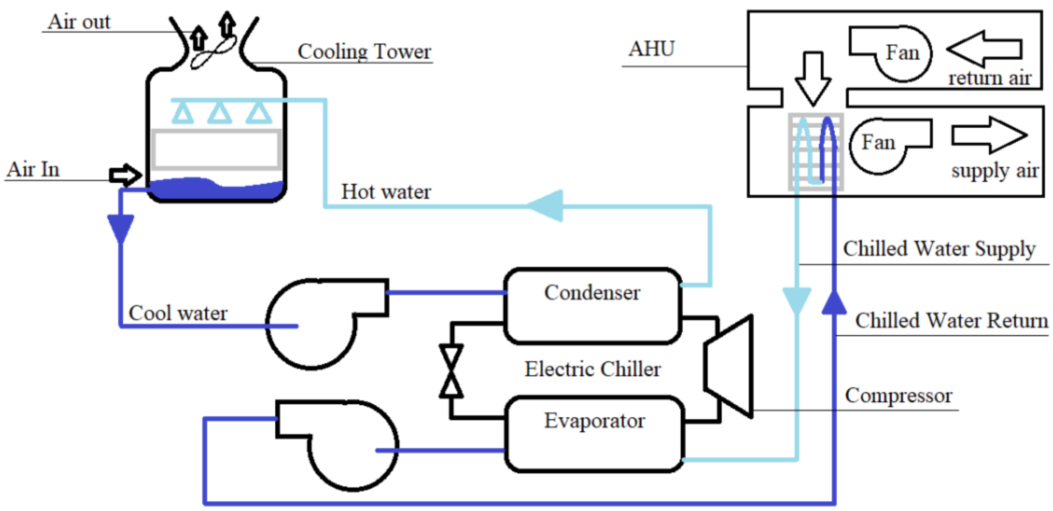


Fig. 5 shows the sequence of Chilled water system operation

2.2 Computer Room Air Conditioner (CRAC) and Computer Room Air Handler (CRAH)

CRACs are one of the oldest cooling systems available. It is inefficient in terms of energy use; however, it is typically affordable. CRAC units are relatively similar to standard air conditioners in that they are driven by a compressor that pulls air over a refrigerant-filled cooling unit. A CRAC's cold air pressurizes the space underneath the elevated floor. The cold air from CRAC enters directly to the front of the server intakes. The hot air is subsequently returned to the CRAC to be cooled after passing through the server rack. Typically, however, this results in mixing with some cold air, lowering efficiency. Often, the return air temperature of the CRAC unit serves as the set thermostat point for the cooling system's operation, resulting in significant temperature disparities between the bottom and top of the server racks. Typically, the CRAC unit's fans operate at a constant speed, and the device has a humidifier that generates steam. Due to the inefficiency of cooling solutions, CRACs should be utilized only with low-density IT equipment.

CRAH units are using the same method as in CRAC units, however, CRAH units are more efficient. A CRAH unit is part of a larger system that includes a chilled water plant (or chiller) located throughout the property. Within the device, chilled water travels via a cooling coil, which is then used to pull air from the outside using modulating fans. Due to the fact that CRAH units function by cooling exterior air, they are significantly more efficient in places with lower annual temperatures.

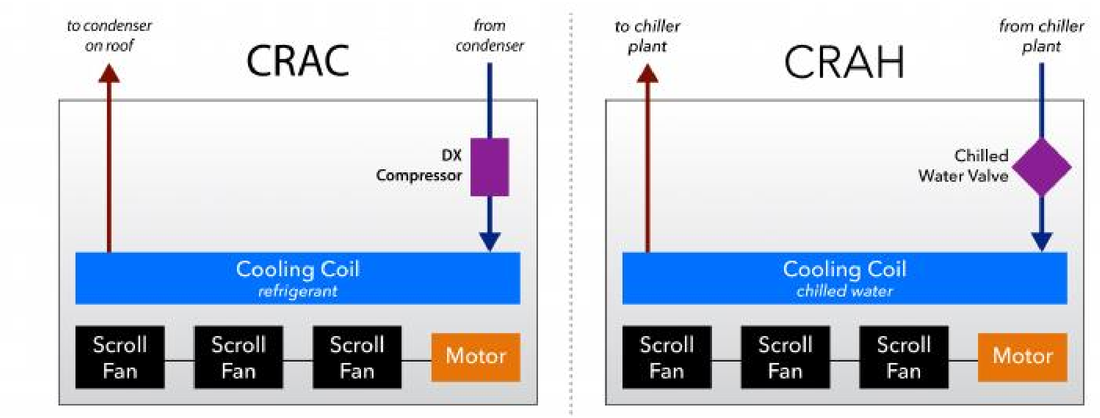


Fig. 6 illustrates the main differences between CRAC and CRAH units (DATA AIRE, 2022)

2.3 Free cooling

Utilizing the cooling system in economizer mode, also referred to as free cooling, is one of the most effective ways to save energy. Free cooling operates on very simple principles. In many parts of the world, the outside air is colder than the temperatures inside a data center for the majority of the year. The usage of free cooling system enables the cooling process to take advantage of externally favourable climate conditions via the by-pass or partial load work of the mechanical active component, resulting in energy and cost savings. As a result, there is no need to cool the hot air in compressors, which implies that the entire process has simply eliminated its most energy-intensive component. That is why free cooling is sweeping the data center industry and rapidly becoming the new standard. Depending on the weather and climatic conditions, the data center's outside air can be significantly colder during certain seasons and times of day than the air warmed by critical equipment inside of the data center. It is possible to use the outside cold air to cool down the servers which are, in turn, reduces the cooling cost. Facebook operates the data center in Lulea, in northern Sweden, just 100 km south of the Arctic Circle where it uses year-round free air cooling. (Informa USA, 2013)

For a data center, there are mainly two primary free cooling options: an inbuilt free cooling or economizer coil on CRAH units or DX air conditioners (indirect free cooling), or a heat exchanger (direct free cooling) that works in conjunction with the chiller. Direct free cooling may be accomplished by sending cold air from the outside directly into the data center when certain circumstances exist outside. Despite its simplicity and efficacy, direct air-side free cooling can bring humidity, particulate matter, and gaseous pollutants into the interior environment, compromising IT dependability. Indirect airside, free cooling, on the other hand, prevents contamination by the use of an air-to-air heat exchanger. Independent free coolers have demonstrated up to a 70% energy reduction compared to mechanical refrigeration. The cost reductions associated with this strategy are a result of the compressor being idle while the fans and pumps continue to operate.

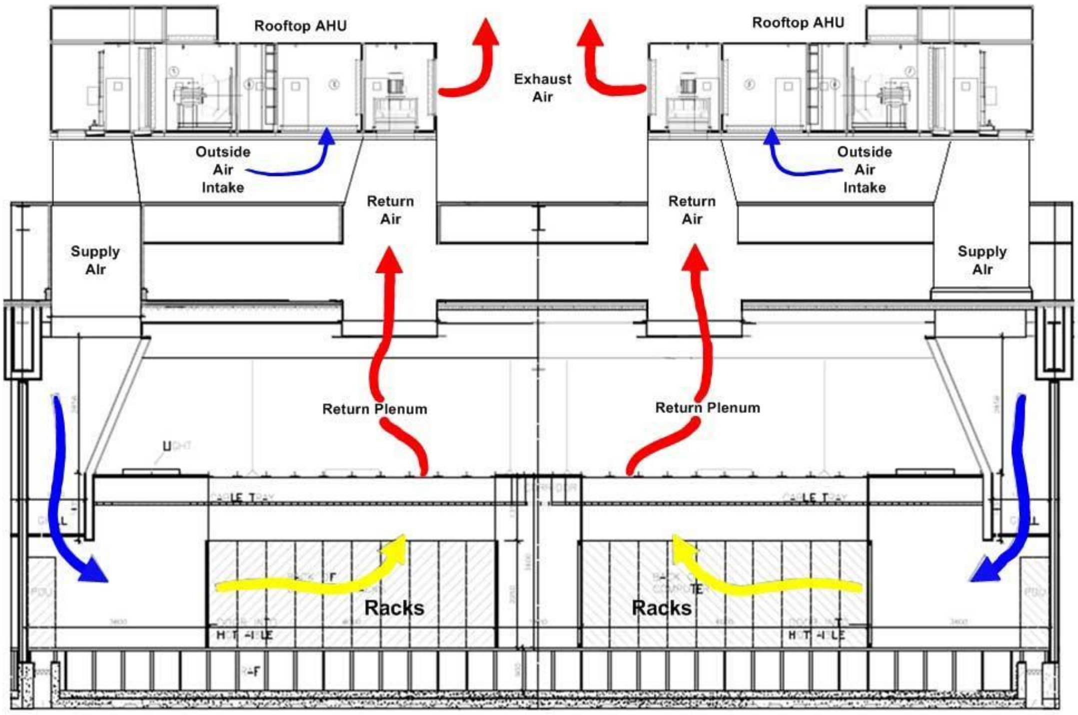


Fig. 7 shows typical Free Cooling Diagram (Informa USA, 2009)

2.4 Cold and Hot Aisle containments

The majority of data centers have used a containment structure that separates the cool airflow from the hot airflow generated by the equipment's exhaust. The containment structure effectively dissipates heat in these racks by isolating hot air from cold air and directing hot exhaust airflow away from equipment and back to air handlers. More significantly, proper airflow management via containment systems enables many cooling system modifications that can result in significant cost savings at any rack density.

Effectively isolated hot and cold air containment systems offer precise control of intake temperatures, appropriate volume adjustment of airflow, and higher return air temperatures, allowing room temperatures to be consistently raised. As a result, cooling system efficiency increases and the number of days those economizers may be utilized for "free cooling" increases. Optimized airflow volume results in the operation of fewer air handlers, increases in chilled water temperatures, which results in enhanced chiller efficiency, and a reduction in power costs due to the overall energy utilized for cooling equipment.

Additionally, complete containment often allows for a reduction in the volume of air given to equipment. Traditional hot aisle/cold aisle systems are frequently oversupplied to compensate for bypass airflow over and around cabinets. Once containment is in place, the volume of airflow may normally be lowered. In older buildings with air handlers that lack variable speed fans, this entails shutting down some units, resulting in some energy savings. In locations with variable speed fans, fan speeds can be decreased to capitalize on fan affinity laws. When fans are operated at reduced speeds, they consume less energy overall since they are moving less air. In essence, all units are operated at a low partial capacity rather than some units operating at full power. This arrangement retains the ability to provide N+1 redundancy or greater if necessary while preserving lower energy costs. It is critical to maintaining a suitable pressure differential between the open and enclosed spaces to direct air toward the air handlers. This difference may be fine-tuned to a negligible level by automation, significantly reducing energy costs.

In conventional open hot/cold aisle layouts, chilled air may be given at a significantly lower temperature, for example, 12.8°C, to avoid supply and return air mixing. However, complete isolation of hot and cold aisles eliminates recirculation of hot return air. Much lower temperature air is no longer required to overcome air mixing, and the supply air temperature may be safely increased to 18-27°C. The benefit of increasing the supply air temperature is increased cooling unit efficiency as a result of the correspondingly higher return air temperatures and more "free cooling" hours when utilizing economizers (Hackenberg) (Chao, Guangming, Shuiquan, & Yueqin, 2017).

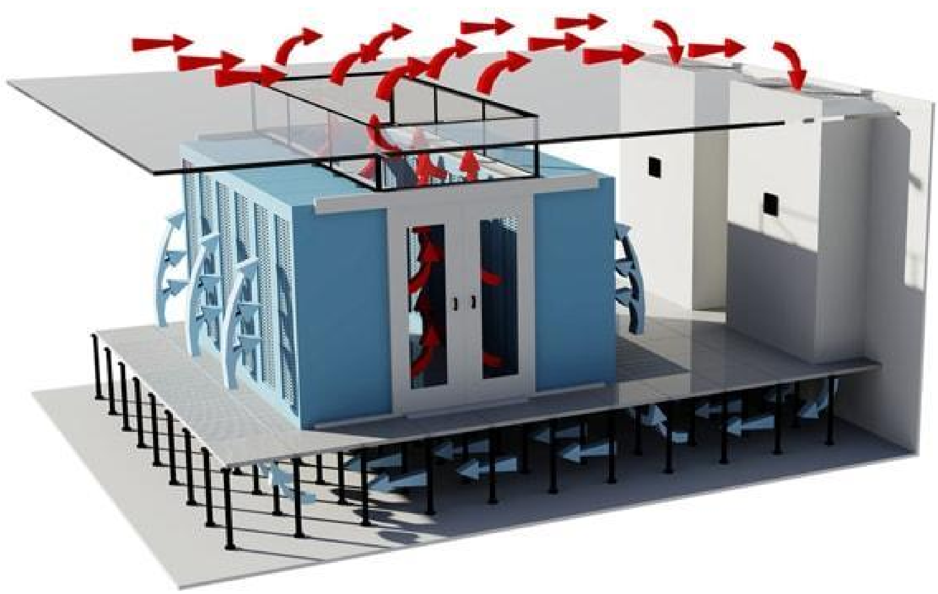


Fig. 8 shows Hot aisle containment arrangement (Colocation America, 2014)

2.5 Raised Floor

Historically, this technique has been utilized in combination with earlier CRAC and CRAH cooling systems. The data center floor is elevated on a frame above the concrete slab level of the structure. This elevated gap between the two is utilized to run cooling water pipes, electrical, and mechanical services or to provide pressured airflow for cooling. In a data center, a raised floor is an elevated floor installed around 60 cm above the concrete floor level. By utilizing a raised floor, facilities may not only minimize the quantity of air required to cool equipment, but also save money and enhance temperature distribution across all cabinets. The data halls with the raised floor might possibly result in a 40 percent less cooling load. (Schiavon, Ho Lee, Bauman, & Webster, 2010)

2.6 Direct to Chip Cooling

When it comes to power density, it is severely constrained in air-cooled data centers due to the inability of air cooling to remove such a large quantity of heat. To achieve efficient cooling, the densities must be extremely low. Efficiency must be sacrificed in order to achieve a larger density in the typical air-cooled data center.

Direct to Chip Cooling (Hybrid cooling) refers to a cooling approach in which elements of the critical equipment that generate a lot of heat are cooled with water while the remaining parts remain air-cooled. Thus, it is a hybrid of air cooling and liquid cooling. Direct to Chip Cooling requires that servers remain installed in standard server racks, just as they are in air-cooled data centers. While server fans continue to circulate air over the equipment, the components that have the potential to get extremely hot are additionally cooled using cold plates. A good example of this type of component is the CPU. Due to the fact that this system directly cools the CPUs, it is one of the most effective techniques of server cooling. Copper or aluminium are used to make cold plates. They feature an entrance and an outlet, which allows water to circulate through them and remove heat without fear of a short circuit. Without cold plates, water could not be utilized to cool electrical equipment due to its excessive electrical conductivity.

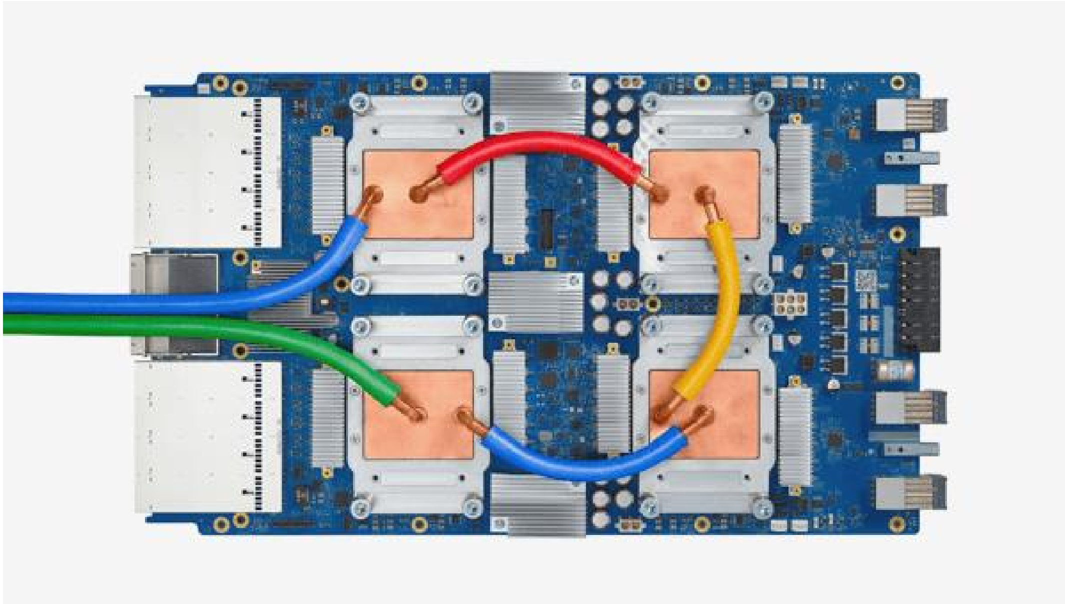


Fig. 9 shows direct to chip cooling principle with four components that are cooled by four cold plates (Google Cloud, 2022)

In this case, the cold plates are composed of metal and are linked by flexible tubing. After absorbing the heat generated by the IT equipment, the water is pushed out of the system and either actively cooled with a chiller or passively chilled with a cooling tower. In a cooling tower, coolant is exposed to the environment indirectly and absorbs heat. Once the water has cooled sufficiently, it can be utilized for cooling again.

This cooling technology is mostly utilized in high performance data centers. Data centers are designed for maximal processing power, with efficiency and cost of hardware being secondary considerations. These systems' high-performance CPUs often have a greater thermal design power, which means they generate more heat than a typical server CPU. Due to the increased heat generated, those data centers are frequently cooled using hybrid cooling. To provide sufficient air conditioning in high-performance data centers, heat sinks would have to be significantly larger, airflow higher, or air colder.

Water has a volumetric heat capacity that is more than 1000 times that of air, therefore Direct to Chip Cooling is far more effective at removing heat from server components than air cooling. The volumetric heat capacity of a substance is the quantity of heat that it can absorb. Hybrid cooling's increased heat capacity also enables more efficient use of waste heat. Another benefit of water's larger heat capacity than air is that it can cool more racks with a considerably higher density with the same amount of coolant. The high-power densities attained with Direct to

Chip Cooling are not attainable with conventional air cooling. Direct to Chip Cooling is also more efficient than air cooling. Hybrid-cooled data centers consume around 10% of its energy for water cooling and around 5% for pumping, leaving up to 85% for IT equipment.

But despite the clear advantages, the system also has major disadvantages which preclude it from being the primary cooling option. One downside of Direct to Chip Cooling is that air-cooling is still required. As a result, many of the issues associated with air cooling remain applicable for hybrid cooling. Corrosion is still a possibility due to the poor quality of the air utilized for cooling. Similarly to air cooling, areas without cold plates might sometimes overheat. Furthermore, due to their integration into the cooling water circuit, it might cause some difficulties in case of server replacement. Special quick-connection points are required, for the hot replacement of the server devices. Hot replacement is a term that refers to the procedure of swapping out one server within a live rack. However, quick-connection points introduce possible failure spots that might result in water leakage. The connecting components of such a system are always the most delicate, for example, if the material is not placed properly, water can leak out between the cooling tube and the cold plate. Finally, the expenditures associated with Direct to Chip Cooling are an unavoidable disadvantage: Due to the fact that Direct to Chip Cooling requires both air cooling and liquid cooling equipment, the initial cost is significantly greater than for air cooling alone. For instance, cold plates, tubing, and fittings are costly.

2.7 Immersion System

Immersion Cooling or also known as liquid submersion cooling is the innovative practice where the entire servers are immersed in the bath with a non-conductive, non-flammable, dielectric liquid with better heat conductivity than air or water. This kind of technology is widely used in the cooling of large transformers; however, it is gradually gaining popularity in the IT industry. Due to the fluid's superior ability to transmit heat more efficiently than air, it is extremely energy efficient and provides excellent cooling performance. Typical liquids used in immersion cooling systems have a heat capacity around 1300 times that of air, which implies that one litre of liquid can transfer approximately 1300 litres of heat. There are two distinct methods of liquid immersion cooling currently available in the market, which will be discussed in detail below.

2.7.1 1-phase immersion

The liquid will not change its physical state and will remain as a liquid during the process of single-phase immersion cooling. The liquid running over the heat generating components cools them down and transports the hot liquid away. There are two ways to circulate the liquid: with the help of the pump or by using natural convection method. Because warm liquid has a greater volume than cooled liquid, it goes to the top in natural convection systems. Then it flows to the tank's side, where it is cooled down by using a heat exchanger which in turn connected to the external loop. The cooled liquid then flows back to the tank's bottom. At the bottom, it is redirected beneath a mainboard and circulated back to heat generating components, where it is re-heated. In the pump driven systems, an electrical pump circulates the liquid via an inlet within the tank and out the other side via an outlet. After that, the liquid is cooled using a heat exchanger to transfer the heat to external the chilled water loop.

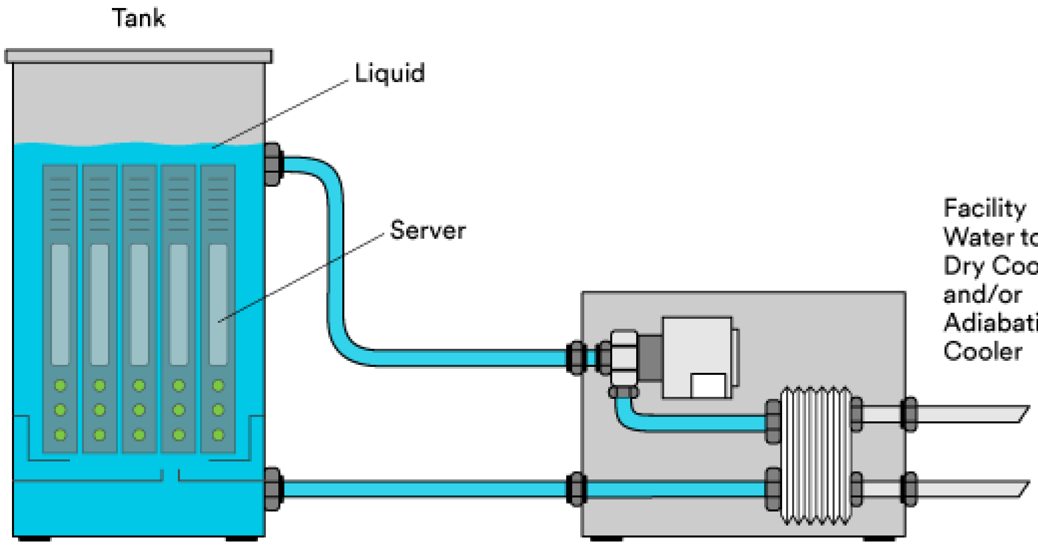


Fig. 10 shows how the liquid circulates in single phase liquid cooling system (3M, 2019)

2.7.2 2-phase liquid immersion cooling system

In comparison with single cooling system, the double phase cooling is distinguished by the fact that coolant changes the physical state when it comes with a contact with a heat generating elements. In order to avoid harming the components, the coolant's boiling temperature must be less than the maximum designed temperature of the elements to be cooled. This process is shown in Fig. 11. The liquid evaporates on the hot elements of the servers, then the gas rises to the top of the tank, making a way for the new, cooler liquid to absorb the heat.

Within the tank, above the liquid, is a condenser. To remove the heat, chilled water is pumped via a condenser. (Ramdas, Rajmane, Chauhan, Misrak, & Agonafer, 2019)

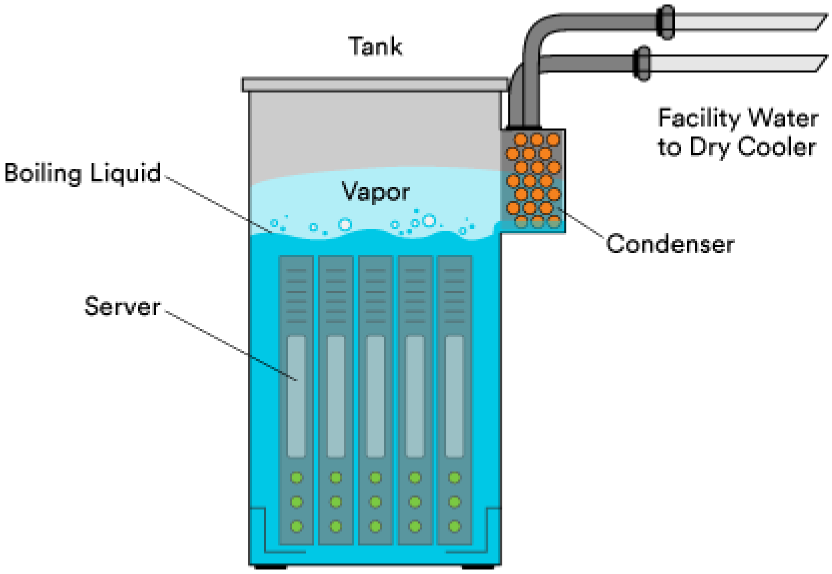


Fig. 11 shows the cooling liquid circulation in double phase immersion cooling system (3M, 2019)

The primary disadvantage of immersion cooling is that it is still in its infancy. Due to the insufficiency of information available on its long-term reliability, data center operators are refrains to install it in their data centers. (Coles & Herrlin, 2016)

Data center operations require further training on proper cooling liquid management. Spills must be cleaned immediately. Additionally, servers must be extracted vertically from the system, which provides a challenge for operations staff because servers can become fairly heavy. This complicates server extraction compared to traditional systems with the CRAC/CRAH units, where servers on metal-framed racks can simply be dragged out horizontally. To keep costs low in a 2-phase liquid system, tanks must be properly sealed or fluid leakage must be avoided.

But if we look at this technology from the other side, we will find that the benefits of an immersion cooling system much exceed the disadvantages. A liquid immersion system absorbs all the heat produced by the server; this results in the highest waste heat utilization of all other cooling technologies mentioned above. (Tuma, 2010)

With this technology, the liquid that is used can be even up to 45°, which is, in turn, enables great free cooling utilization. Due to the extremely consistent temperature distribution, overheating of elements that would be exposed to heatsinks is avoided. When compared to the energy required for air cooling, the fluid used in a single-phase system is less expensive. Additionally, liquid immersion requires extremely few mechanical components, just a pump to circulate the liquid. The elimination of mechanical parts increases the servers availability and the entire cooling system reliability.

2.8 The benefits and drawbacks of various cooling methods.

Most of the cooling techniques are used in the data center industry mainly focused on power density and usage efficiency. Historically, the majority of data centers have cooled their rooms using air cooling technologies. This sort of solution is built on hot aisle and cold aisle tactics. (Lu, Zhang, & Yang, 2018) It entails aligning the data center's racks in such a way that they face alternately cold and hot air intakes and exhausts, so producing cold and hot rows.

One of the most efficient air techniques, on the other hand, is free cooling (Daraghmeh & Wang, 2016), which takes use of the favorable climatology of cold areas to cool the data center's IT equipment. Cooling expenses decrease significantly as a result of the removal of much of the unnecessary free cooling equipment. Large Cloud services providers place their data centers in favorable to free cooling locations, reaching remarkable power usage efficiency (PUE) factors (Meta, 2020), as PUE is ratio between total load and IT load (useful). However, studies indicate that free cooling methods are prone to failure and cannot be used in all areas. (Heinemeier, 2017)

If we review separately only air-cooled system disadvantages, the one of the biggest disadvantages is that the air has lowest heat transfer coefficient to compare with all other cooling techniques. (Nadjahi, Louahlia, & Lemasson, 2018) Additionally, corrosion, vibration, and turbulence are all caused by air fluxes. As a result, possible failures exist.

Water based cooling technique address these issues and provides superior heat transfer characteristics, resulting in increased energy savings. Around 50% of energy used for the cooling of IT equipment can be saved as compared to an air-cooling technique. (Nadjahi, Louahlia, & Lemasson, 2018) Nevertheless, with this kind advantage, it has one significant flaw: leakage which have the potential to be fatal for the servers and network devices located in data centers.

Fig. 12 shows the typical liquid cooling rear door heat exchanger (Loyolan Ventures, 2021)

The rear door cooling system, which is available in passive and active configurations, is one of the most essential water-based systems. It utilizes heat exchangers but relies on the fans on the IT equipment to propel the airflow. They are capable of handling up to 30kW per rack in perfect conditions, however it may be more cost effective to strive for lower levels. (Karki, Novotny, Radmehr, & Patankar, 2011) Direct-to-chip water cooling has the highest power density per rack of all water-cooling methods, reaching sometimes up to 80kW in certain commercial devices. (Chainer, Schultz, Parida, & Gaynes, 2017) Additionally, it operates at greater working temperatures, which reduces energy expenditures. The issue is that it requires specialized hardware, which restricts its impact. Dielectric-liquid cooling is gaining popularity due to its extraordinary potential. The solutions are numerous. The indirect dielectric cooling solution avoids direct contact with the equipment and similar to certain water solutions. (Li, Zhou, Tian, & Li, 2021) Unlike such tactics, leakages are no longer a common thread. However, in case of indirect cooling, this system has less heat removal capacity rather than in direct liquids since it is not in touch with them. Another downside, like with liquid-based remedies, is that it is equipment-specific. As a result, it is frequently unable to be repurposed.

Additionally, the equipment's architecture, which is concentrated on several pipes, makes manipulation difficult.

Immersion Cooling is the most energy-efficient in dielectric liquid cooling techniques. In single-phase immersion cooling, the equipment is immersed in an open bath. (Day, Lin, & Bunger, 2019) This approach takes use of the liquid's heat transfer potential but requires a large recirculation system to cool it down, reducing energy and space efficiency.

There is an increased cost in immersive cooling systems due to the fluids as it requires special mineral oils and engineering fluids. (3M, 2008) In the case of mineral oil, despite its wide availability in the market, it has many disadvantages such as high Global Warming Potential (GWP) as only around 35% of the entire oil is biodegradable and it is very toxic. (United States Environmental Protection Agency, 2021) However, the engineered fluid is a special non-flammable fluid that is specially designed to use in heat transfer applications. Additionally, it is non-toxic, have a low GWP index and almost 90% biodegradable, and have a greater heat transmission coefficient. Usually in 2-phase-immersive systems, used the liquids with a low boiling point, since electronic equipment submerged in a closed bath evaporates the coolant during operation. In a 1-phase system, the liquids with higher boiling points are used.

Considering the prospects for the development of all the technologies we have considered above, then it turns out that the 2-phase immersion cooling technology has a great future, since it has the potential to cut cooling energy usage by up to 95% While air-based systems typically have a power density of the servers of 4-40 kW, however in case of double immersion cooling the density can reach up to 250 kW per rack alongside with the PUE around 1.02-1.03 which was obtained only before by free cooling systems.

Boiling Point °C	3M™ Novec™ Engineered Fluids					
	products for two-phase liquid cooling applications			products for single-phase liquid cooling applications		
	7000	7100	7200	7300	7500	7700
	34	61	76	98	128	167

Table 4 illustrates the boiling points of the different 3M Novec Engineered Fluids (3M, 2020)

Cooling System	Advantage	Disadvantage
Air cooling	The most widespread system is capable of utilizing the outside air.	Energy and space expenses are high, and they are location dependent.
Free cooling	External heat exchanger. Significant energy savings.	For safety, it requires a acceptable climate and an additional cooling system.
Hot/Cold aisle containment	Distribution of distinct hot and cool airflows inside an aisle	Susceptible to unfavorable hotspots and turbulence.
Water based cooling system	Eliminates air cooling concerns such as rust and vibration. Coefficient of heat transmission increased.	Coefficient of heat transmission is lower than other liquid-based systems. Leakages can be fatal.
Dielectric liquid cooling	Significantly safer and more energy-efficient than water systems.	Currently under development and not frequently used techniques.
1-phase immersion cooling	Potential for heat transmission in open bathtubs.	Requires low coolant temperatures when connected to an external circuit. Under development
2-phase immersion cooling	Passive cooling approach that conserves up to 95% of cooling energy.	High pressures that may exist in enclosed bathtubs. Under development.

Table 5 illustrates the main advantages and disadvantages of the different cooling systems.

2.9 Summary

This chapter describes the various and most common types of cooling systems available on the market since the cooling system is the largest consumer of electricity after the IT equipment in the data center. The expensive cost of cooling equipment is one of the primary reasons enterprises shift away from on-premises data centers and toward colocation. When it comes to cooling IT equipment, the majority of private data centers are highly inefficient. Additionally, they lack the monitoring capabilities of colocation data centers, making it more difficult to fully optimize infrastructure for cooling demand reduction. It is the optimization of the cooling system that will have the greatest positive effect in terms of energy efficiency. That is not all. Inefficient cooling systems can drastically raise operational expenses. Also in this chapter, the advantages and disadvantages of various cooling systems are considered in detail. Special attention is paid to the liquid immersion cooling technology. This innovative knowledge and characterization of the thermal generating and harvesting processes allows the computer-aided design of dynamic flow provisioning methods with the goal of either cost-effective operation or improved heat harvest quality.

3. Efficiency differences between air and liquid cooling.

From a computer standpoint, it is ideal to have a figure that approximates computational capacity in proportion to energy consumption. Historically, computing performance has been quantified by performing a certain number of operations on a computer and recording the time required to complete them; this is referred to as a benchmark. FLOPS is the number of computational operations or instructions performed per second on floating-point operands (FP). The term "computational" refers to the fact that the microprocessor may also perform other instructions, such as loading from memory as they do not impose a beneficial computing load and are thus omitted. The FLOPS number is characteristic of the CPU itself and not of the program. The open benchmark from “Lapack” is used to calculate the FLOPS. (Netlib , 2021) The calculation's outcome is expressed in floating-point operations per second (FLOPS) (1).

$$computing\ performance\ (FLOPS) = \frac{floating\ point\ operations}{time}$$

(1) This formula is to calculate Computing Performance.

If we divide the benchmark value by the system’s peak power we will get the value that can help to us to compare the efficiency of other systems, as this will represent the FLOPS/W value which may be used to compare the performance per to the amount of energy required. In the below formula (2) and also further, the variable μ will show the value of various computing efficiencies.

$$\mu\left(\frac{FLOPS}{W}\right) = \frac{computing\ performance}{maximum\ energy}$$

(2) This formula is to calculate the server’s computing efficiency.

PUE – Power usage effectiveness.

Green Grid is a non-profit organization comprised of businesses, universities, and organizations formed in 2006 with the mission of increasing the efficiency and sustainability of data centers. (The Grin Grid, 2022) To this aim, "The Green Grid" has developed a variety of measures, including PUE, CUE, and WUE, to assure worldwide comparability of measurements and resultant conclusions. One of the most critical measures in this context is PUE (Power Usage Effectiveness) which is defined as follows (The Green Grid, 2012):

$$PUE = \frac{Total\ Facility\ Power}{IT\ Equipment\ Power}$$

(3) This formula is to calculate the power usage efficiency.

Where:

IT equipment power encompasses the load generated by all types of IT equipment, including computers, storage, and network equipment, as well as auxiliary equipment. The facility's entire power output is comprised of the following:

- Power supply components such as UPSs, switchgear, generators, Power Distribution Units (PDUs), and batteries, as well as distribution losses external to the IT equipment.
- Components of the cooling system, including chillers, Computer Room Air Handlers, pumps, motors, and etc.
- IT equipment power.

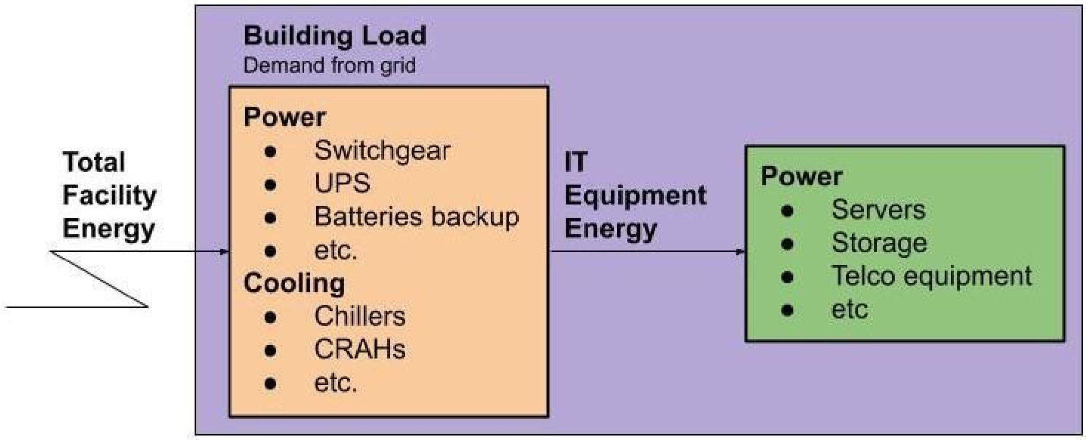


Fig. 13 – total facility energy and IT equipment energy

To determine the efficiency differences between immersion and air-cooling data centers, PUE values must be transformed. The inverse of value PUE which is equal to IT load divided by total power represents the percentage of energy consumed by IT equipment. For example, the data center with the PUE equal to 2, 50% of the total energy used in this data center, consumed by IT equipment. The inverse of the PUE value can also be used to calculate FLOPS/W to calculate the efficiency increase associated with the switching to the immersion cooling technique (4).

$$proportion\ of\ the\ IT\ equipment\ energy = \frac{1}{PUE} = \frac{Energy\ consumed\ by\ IT\ equipment}{Total\ facility\ power}$$

(4) This formula is to calculate the proportion of IT energy in the total energy used in the Data center.

PUEs in data centers with air cooling systems range between 1.1 and 2.9. Values like 1.1 are only possible in hyper-scale data centers that have been carefully designed for energy efficiency. For example, Google's data centers achieve a PUE of 1.10. This indicates that 89% of the energy consumed by Google's servers is used for computation. However, energy-efficient data centers account for a negligible portion of the energy consumed by data centers globally. In comparison to the most efficient centers, average air-cooled data centers have a much higher PUE. The average PUE ratio for data centers is 1.58 and in the case of liquid immersion cooling data centers the average PUE ratio is 1.03 which means about 97% of the energy consumed by IT equipment (Lawrence, 2020).

Changing the cooling system has no effect on the processing performance or the energy consumption of the IT equipment. By replacing the cooling system, the only value that changes is the total energy consumed by the equipment. As the IT energy and FLOPS values are the constant values, the designed thermal power of the IT equipment will remain the same regardless of the used cooling system. The energy used by IT equipment is E_{it} , b_o is floating point operations, b_t is time required to complete the floating point operands (5).

$$\mu = \frac{\frac{b_o}{b_t}}{E_{it} \times PUE} = \frac{\frac{b_o}{b_t}}{E_{it}} \times \frac{1}{PUE}$$

(5) This formula shows the relation of computing efficiency to the total energy used by IT equipment.

As it was already mentioned above b_o , b_t and E_{it} are the constant values, we can deduce from this formula that the computing efficiency or the relation of the computer performance to the energy used only depends on the part of the energy used for the IT equipment. In order to measure the changes in the performance that we can gain by switching to the immersive cooling system, we must take the PUE values from both systems, immersive and air-cooled systems (6).

$$\frac{\frac{\frac{b_o}{b_t}}{E_{it}} \times \frac{1}{PUE2}}{\frac{\frac{b_o}{b_t}}{E_{it}} \times \frac{1}{PUE1}} = \frac{\frac{1}{PUE2}}{\frac{1}{PUE1}} = \frac{PUE1}{PUE2}$$

(6) This formula shows that in order to measure the performance of IT equipment, the PUE values can be used from two different cooling systems.

The best PUE value for the air-cooled system is equal to 1.10, however the average PUE value from immersion cooling system is 1.03, by simple calculation we can get 106,79% of efficiency, this means that the efficiency that can be gained by switching to the immersion cooling system will be around 6,79%. The same calculation can be done by using the average PUE value and immersive cooling system PUE value: 1.58/1.03=153,39%. The improvement of efficiency is 53,4%.

This formula can be applied also to the sample processor used in servers, to show that there is no dependency between the CPU and efficiency value. As an example we can use one of the best workstation CPUs available on the market: AMD Ryzen Threadripper PRO 3995WX which has 12,902.4 GFLOPS with a maximum consumption of 280 W (Tech Power Up, 2022) (7).

$$\mu_{AMD\ CPU} = \frac{12{,}902.4}{280} \approx 46.08\ GFLOPS/W$$

(7) Calculation of the performance for AMD CPU

This is the maximum efficiency of the processor, however, the power consumed by IT equipment does not only used by the server processor that calculates the FLOPS, but it is also necessary to find the power CPU's power consumption in comparison to the whole equipment. The power consumption of the processor, relative to the total power consumption of the IT equipment is around 80% (Armenta, et al., 2015) (Gough, Steiner, & Saunders, 2015) (Zhang, Lu, & Qin, 2013). Knowing the proportion of the CPU's power, we can calculate the efficiency value for the entire IT equipment (8):

$$\mu_{IT\ equipment} = \frac{12{,}902.4}{280} \times 0.8 = 36.86\ GFLOPS/W$$

(8) The performance of entire IT equipment

By using the PUE values for different cooling methods, we can now calculate the efficiency value (FLOPS/W) for the data center (9). This allows us to compare the efficiency of immersion cooling and air-cooled data centers by using constant values.

$$\mu_{immersion\ cooling} = \frac{12{,}902.4}{280} \times 0.8 \times \frac{1}{1.03} = 35.79\ GFLOPS/W$$

$$\mu_{best\ practice\ air\ cooling} = \frac{12{,}902.4}{280} \times 0.8 \times \frac{1}{1.10} = 33.51\ GFLOPS/W$$

$$\mu_{average\ air\ cooling} = \frac{12{,}902.4}{280} \times 0.8 \times \frac{1}{1.58} = 23.33\ GFLOPS/W$$

(9) The performance of IT equipment for different cooling methods

To demonstrate the effect of switching to immersion cooling to efficiency, all numbers have been converted to FLOPS/W, however, errors introduced by this conversion have no effect on the conclusion of this evaluation as all outcomes are similarly affected. Comparing the results of the calculations, as can be seen, a significant improvement in the efficiency can be gained by switching to immersion cooling.

3.1 Summary

This chapter compares the two most important cooling systems. It has been mathematically proven that performance directly depends on the type of cooling system and the best result can be achieved using immersion cooling system. Several categories pertinent to data center operators were found during the research on liquid immersion cooling. The key areas are overall data center efficiency, data center density with all of its supporting equipment, and data center reliability. Additionally, waste heat use and costs are significant considerations. As demonstrated in this chapter, data centers cooled through liquid immersion cooling have been proved to be significantly more efficient than those cooled via air. The results demonstrate a gain in efficiency associated with the conversion of a typical air-cooled data center to immersion cooling. Immersion cooling can double the efficiency of data centers, and even the most efficient air-cooled data centers can improve their performance by roughly 10%.

4. Cooling system optimization.

The most often used data center design structure is raised floor. There are supply and return pathways in every air distribution system. The following are the three basic methods for transporting air between the CRAH and the IT equipment:

• Flooded

• Ducted Locally

• Ducted Fully

Flooded, Locally Ducted, and Fully Ducted are the three ways that can be utilized in either the supply or return paths. As a result, there are nine different types of distribution systems that can be created. All these types have been utilized in a variety of situations, and multiple types are sometimes blended in the same data center. Some of these techniques necessitate a raised floor, while others can be utilized with either a hard or raised floor (Rasmussen, 2005).

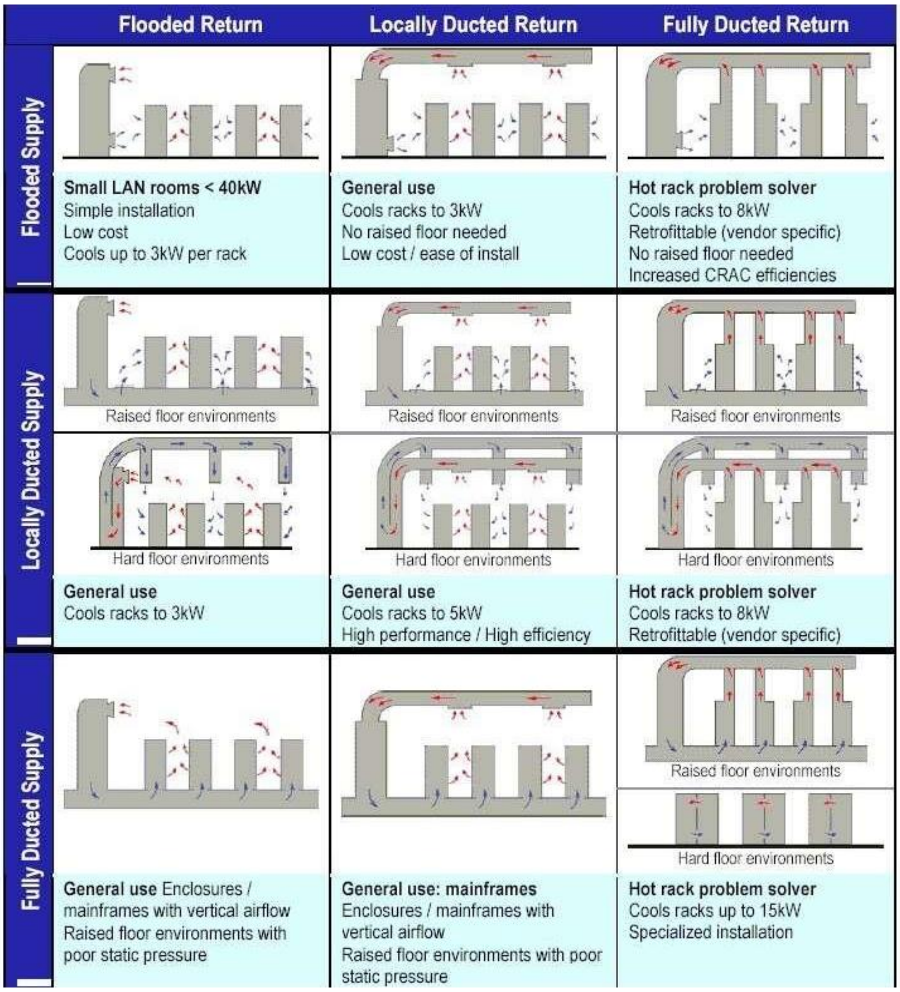


Fig. 14 illustrates the 9 types of cooling systems (Rasmussen, 2005).

A cooling system's objective is to maintain an appropriate temperature and humidity range for rack inlets via air distribution systems. The intake conditions are now referred to in most ecological requirements. The most recent Thermal guidelines for Data Processing Environments were published by ASHRAE in 2015, as seen in Table 6. In table 6 demonstrates that the temperature of intake air entering IT equipment should be kept between 18 and 27 degrees Celsius, with a minimum dew point of 5.5 degrees Celsius and a maximum dew point of 15 degrees Celsius.

Temperature control range			Moisture control range
Class	Dry-bulb temperature ˚C	Recommended level ˚C	Maximum Dew Point ˚C
A1	15-32	18-27	17
A2	10-35		21

Table 6 shows thermal guidelines for Data Processing Environments (ASHRAE, 2021)

4.1 Air management and challenges

Air management's objective is to keep critical (or IT) equipment's intake conditions within prescribed limits (e.g. ASHRAE) while consuming the least amount of energy possible. The desired method of air management is to supply cold air as close as feasible to the IT equipment intakes, with the IT equipment inlet temperature being equal to or close to the CRAH units cold supply air temperature. As a result, the cold supply air temperature can be increased close to the maximum recommended value (e.g., 27C, Table 6) to optimize chiller energy efficiency. Raising the supply air temperature of the CRAH unit enables the chillers to operate at a higher supply water temperature while utilizing more free-colling and less compressor mode (i.e. chilled water temperature), resulting in increased energy efficiency. The energy savings from increasing the cold supply air temperature alone are estimated to be in the 15–25 percent range. Additionally, proper air management enables the CRAH units' fans to run at a lower speed and utilize more cold water from the chillers, thereby conserving fan power.

Using Hot aisle and cold aisle, it is feasible for hot exhaust air to recirculate and mix with cool air at the rack input. The combination of hot exhaust air and cool supply air may result in an increase in the inlet temperature and the formation of hot spots which are in turn (rack inlet conditions that are excessively hot or dry) one of main problems in data centers. According to a recent survey from Uptime Institute (Sullivan, Strong, & Brill, 2007), approximately one in ten racks operates at a higher temperature than specified reliability criteria (e.g. ASHRAE), and the majority of hot spots occur in data halls with low load. This research reveals that the root cause of hot patches is not a lack of cooling capacity or a high heat density, but rather a lack of proper air management. Reducing bypass airflow to 10% or less for the sites can result in a 2°C to 4°C reductions in cold aisle temperature while enabling the cooling unit's return air temperature control setting to be increased. In terms of air management, two significant issues are recognized and linked to the present Hot aisle and cold aisle containments installed in the data hall: bypass air (Fig. 15b) and recirculation air (Fig. 15a).

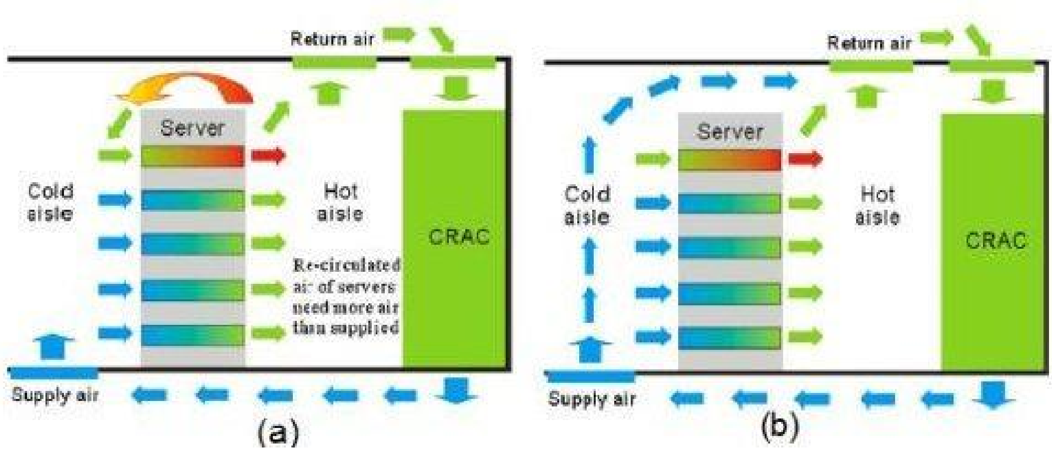


Fig. 15 shows how the Data halls air circulation flow. (a) recirculation airflow and (b) bypass airflow (Nada, Elfeky, Attia, & Alshaer, 2015).

Air bypassed: Bypass air is not used to cool the equipment, which should be kept to a minimum. The causes might be an overabundance of supply air or cable cutoff leakage. Excessive bypass airflow rate has been recognized as a root cause of inefficient cooling and hot spots. Numerous studies (Sullivan, Strong, & Brill, 2007) indicate that around 59% of the cold air from the CRAC unit bypasses the computer equipment's air intakes, leaving just 41% of the cold air cooling the equipment directly. With so little cool air entering computer equipment's intakes, heat removal is really accomplished by a mixture of bypass air and hot exhaust air, which introduces another issue: recirculation air. Additionally, research indicates that hot areas can be avoided by lowering the rate of bypass airflow (Sullivan, Strong, &

Brill, 2007). On the other side, recirculation air contributes to cooling equipment numerous times, which should also be reduced. The cause might be a shortage of supply air. Recirculation air frequently results in an increase in the equipment's intake temperature, which is one of the primary causes of hot spots. Several techniques and performance indicators for evaluating and improving air management have been recommended in different studies (Herrlin, Airflow and Cooling Performance of Data Centers: Two Performance metrics, 2008) (Bash, Patel, & Sharma, 2011) (Bash, Patel, & Sharma, 2006).

The next section will provide an overview of various performance indicators.

4.2 Cooling efficiency metrics

To begin, I'd like to explain three performance indicators and two critical equations that will be used in this study.

4.2.1 Return temperature and supply heat indexes

Metrics of performance: Performance metrics are primarily used to assess the air management system's performance. RTI (Return Temperature Index) is one of the key performance indicators (KPI) that is used to determine the actual usage of available airflow (Herrlin, Data Center Air Management Research, 2010). RTI is defined as

$$RTI(\%) = [\frac{T\ return - T\ supply}{\Delta T\ IT\ equip}] \times 100$$

(10) This formula is to calculate the Return temperature index

$T\ return$ is the temperature of the return air for CRAH unit and $T\ supply$ is the temperature of the supply air for CRAH unit. The $\Delta T\ IT\ equip$ refers to the temperature increases that occur throughout IT equipment.

Table 7 - Return Temperature Index metric

Benchmark value	RTI %
Ideal	100%
Superior	95%-105%
Recirculation	$> 100\%$
Bypass	$< 100\%$

If RTI value higher than 100%, indicates of the air recirculation which increases the return temperature and below 100% indicates of the bypass air in the hot and cold aisle containment which bypasses the rack and directly returns to CRAH unit reducing the return temperature.

The other important performance indicator that is also used to evaluate the air management's performance is SHI (Supply Heat Index). The purpose of SHI is to measure the magnitude of hot and cold aisle air mix which is defined as below (Jin, Bai, An, Ni, & Shen, 2020):

$$SHI = \frac{T\ inlet - Tsupply}{T\ outlet - T\ supply}$$

(11) This formula is to calculate the supply heat index.

$T\ inlet$ is the rack inlet airflow temperature; $T\ outlet$ is the airflow temperature exhaust from rack; $T\ supply$ – is the airflow supply from CRAH unit

Table 8 - Supply Heat Index metric

Benchmark value	SHI
Ideal	0
Superior	< 0.4

Supply Heat Index values are typically less than 0.4, and the lower the number, the better. Supply Heat Index values is solely dependent on the temperature of the air entering the rack, exiting the rack, and exiting the CRAC. Additionally, this value may be computed for a single rack to detect local hotspots. Increased $T\ supply$ implies an increase in entropy due to air mixing, and therefore Supply Heat Index may be regarded a metric for energy efficiency. A high Supply Heat Index value frequently indicates mostly recirculation air, but also indicates that the rack is supplied with air more than it is really needed. In perfectly adjusted cooling system, all Supply Heat Index values should be near to one another regardless of rack heat loads, ensuring that hot and cold air do not mix and that the rack with the highest heat load receives more airflow. A cooling system with very unequal Supply Heat Index value shows that the cold supply air is not adequately dispersed.

4.2.2 Power usage efficiency

The PUE value is computed as the ratio of a data center's overall power consumption to the calculated power consumed by active components such as servers, storage, and network. The less energy used by the IT equipment (for example, climate control, power protection, and lighting), the closer the PUE is to "1". This theoretical number is achievable only if the whole data center's current is consumed exclusively by servers, storage, and other active IT components. A PUE score does not reflect a data center's overall energy usage. As a result, the PUE alone does not tell us whether or not power is indeed conserved.

The fraction of power required for cooling reduces when the server inlet temperature is increased. However, if the temperature rises to the point where the servers' fans operate at maximum speed to offer emergency cooling, the IT equipment's power consumption would increase. While the PUE appears to be better, the real power usage increases.

According to German Digital Association Bitkom, with a physical utilization of the available rack space of approx. 50%, a well-planned, newly built and well-operated data center should be at least 1.4 or better (Bitkom, 2011)

Energy balance equation:

Total heat from IT equipment = Total Cooling power from CRAH units

The following equation may be used to calculate the total cooling power of CRAHs:

$$P = Q * \rho * c * \Delta T$$

(12) This formula is to calculate the total cooling power.

From this formula P is equal to the cooling power; the airflow rate is Q ; air density is – ρ; c is the specific heat capacity of air which is equal ≈1.005 kJ/kg at 300 K (26,85 °C) (Urieli)

ΔT is the difference between return and supply air on the CRAH unit.

The cooling capacity can also be calculated based on the chilled water system but in this case, instead of air density, needs to be used the water density and specific heat capacity of the water. Apart from IT load, the UPS systems, Transformers and PDUs also the units that contribute to the heat load.

4.3 Cooling system efficiency – detailed study

A highly efficient air-cooling system would feature effective air management, allowing for maximum utilization of chilled air with minimal or no bypass or recirculation air. The examination of an air-cooling system should ideally include the calculation of some performance metrics using measurements to determine the cooling, power, and chilled air consumption. Additionally, there is always a requirement for a favorable temperature environment for IT equipment.

The Data Center is in Frankfurt am Main, in Germany. As Frankfurt is one of the most important financial centers in Europe therefor the main function of the data center in our case study is to serve the financial service companies and organizations. The data halls design structure in this data center is a raised floor with racks grouped in a Hot Aisle and Cold Aisle. The CRAH unit cools the return air and then pumps it back to the racks through the raised-floor plenum. The data center also has an outside air economizer with a chilled water-cooling system. One data hall is designed for 2.4MW which is equal to 690 tons of cooling capacity.

4.3.1 Structure and cooling system

The data halls have an open halls approach, and the typical standard room is designed with aisles that contains 30 racks on either side. A typical rack dimension is 610mm (width) x 1,220mm (depth) x 2,300mm (height). The examination space follows the typical Hot Aisle/Cold Aisle design, with locally ducted supply and flooded return. Ther data hall fully ducted supply cooling method separates hot and cold air perfectly, eliminating hot spots and increasing cooling effectiveness. Detailed characteristics of the IT equipment and cooling and electrical system are summarized below:

Designed IT capacity:

- Designed capacity is at 2.4 MW
- 1350 kW is the actual load

Cooling system:

- The locally ducted supply path with a flooded return air distribution

- 16 x CRAH units with N+1 redundancy: 15 of 16 CRACs ON @ 32,050 CMH (18,864 CFM) each with Supply Air Temperature Set Point at 22°C.
- Chilled water delivery pump:
- 2 x Chillers with air-side economizer/free cooling mode that turns on when the outside temperature goes below 10-15 ◦C
- Data Hall Cold Aisle Temperate: 27.0 °C.
  - Cold aisle temperature refers to rack inlet temperature at 1.5m above-raised floor.
  - Supply air temperature reset between 27.0 °C and 32.3 °C to maintain supply air RH below 80% as the outside wet-bulb temperature exceeds 25.6 °C.
  - The mechanical design has complied with the requirements of ASHRAE Environmental Class 2. (Table 6)

    Normal Operating Server Inlet Temperature Limits: 10 – 35 °C.

4.3.2 Computational Fluid Dynamics (CFD) analysis

Even at the planning stage, a detailed computational fluid dynamics (CFD) analysis was made of the data center in order to analyze the fluid flow and compare the airflow properties of the data hall before and after Cold Aisle Containment installation. Computational Fluid Dynamics is the technique of mathematical modelling and numerically solving a physical phenomenon involving fluid flow.

Model #1 is the numerical simulation model that was conducted using a comprehensive CFD modelling model of the air distribution system of the data hall without cold aisle containment.

Table 9 below shows Model 1 without containment

Total Airflow Demand	435,393 CMH
Total Airflow Supply	480,750 CMH
Total Estimated Heat load	2,438 kW
Range of Max Inlet Temperatures	
Above 27°C	195 racks

Between 24 °C and 27 °C	130 racks
Between 21 °C and 24 °C	0

A three-dimensional (3D) virtual white space was created for a reference data center's data hall. The floor layout of the target space for the CFD study is depicted in Fig. 16.

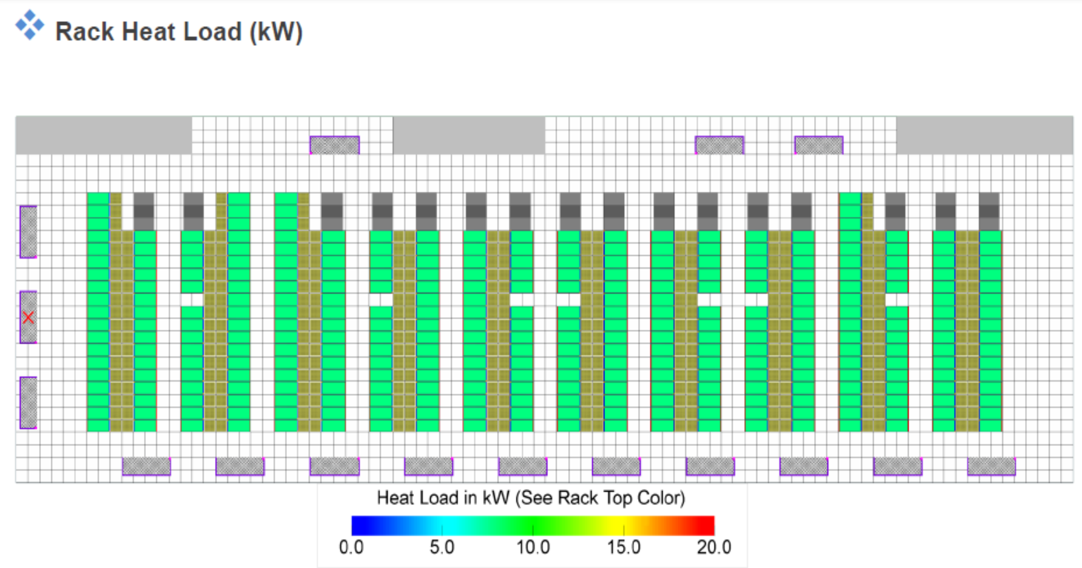


Fig. 16 illustrates the data hall's floor layout with rack load heatmap

Fig. 17 and fig. 18 for Model #1 shows that due to the absence of containment separation of the hot and cold aisles, many servers with the "Red dots" have the inlet temperature above 27 °C and the hot exhaust air mixing with the cold supply air at the cold aisle. There were 195 racks with server-side inlet temperatures above ASHRAE's recommended maximum of 27°C. Several racks had server-side inlet temperatures above ASHRAE's allowable maximum of 32°C due to hot exhaust air recirculating back into the cold aisles.

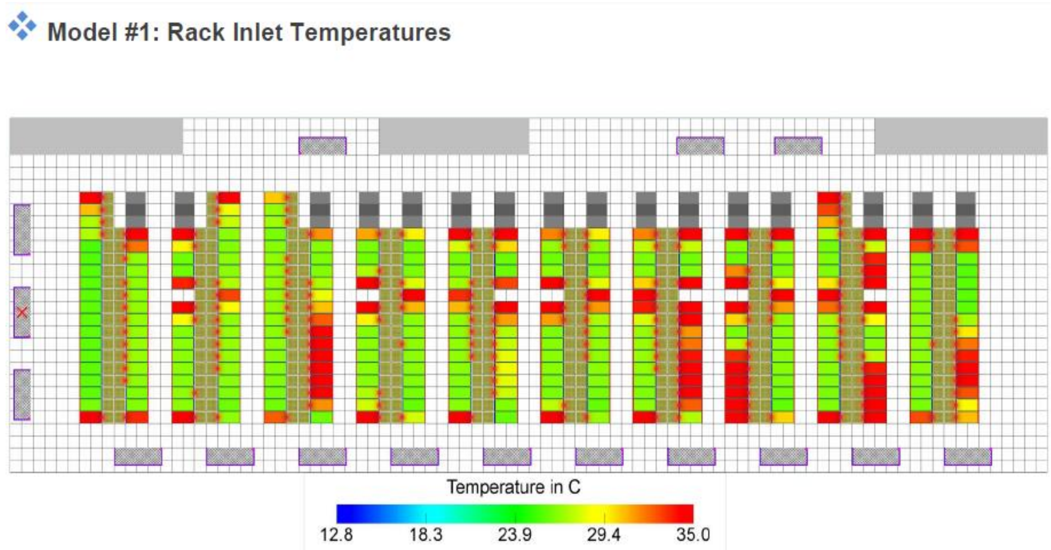


Fig. 17- data hall's floor layout – hot spots map

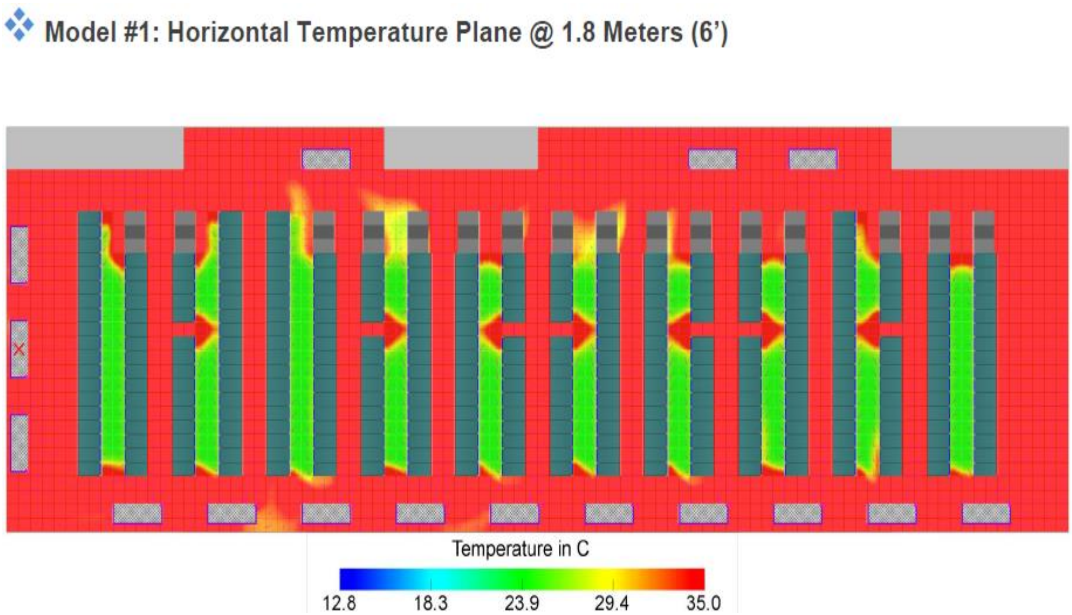


Fig. 18 illustrates data hall's floor layout with hotspots on the rack levels and hot air recirculation.

The below (Fig. 19) supplemental model (Model #2) was used to analyze the effect of installing full Cold Aisle Containment (CAC) featuring doors on the end of the cold aisles and lids on top of the cold aisles.

❖ **Model #2: 3D Floor Layout**

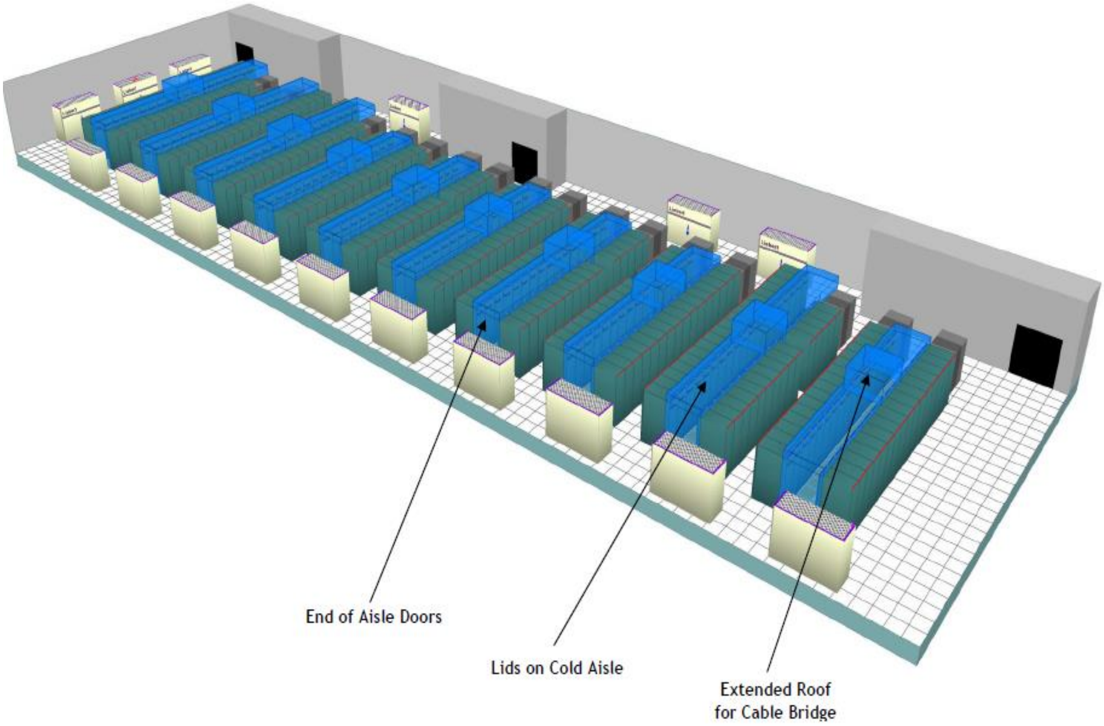


Fig. 19 shows data halls 3D floor layout

With containment installed, there were 0 racks with server-side inlet temperatures above ASHRAE's recommended maximum of 27°C as is shown in Fig. 20. The air distribution for most of the cold aisles was maintained at approximately 26 °C. This is less than 1 °C higher than the supply air temperature of the CRAH units (25 °C), which indicates that the temperature rise was under excellent control. This indicates that the air-cooled in the CRAH units were evenly distributed, resulting in little heat loss or, in other words, the air recirculation was well prevented as originally intended by the cold aisle containment. The effectiveness of the cold aisle containment in terms of prevention of hot air mixing with cold supply air is shown in Fig. 21.

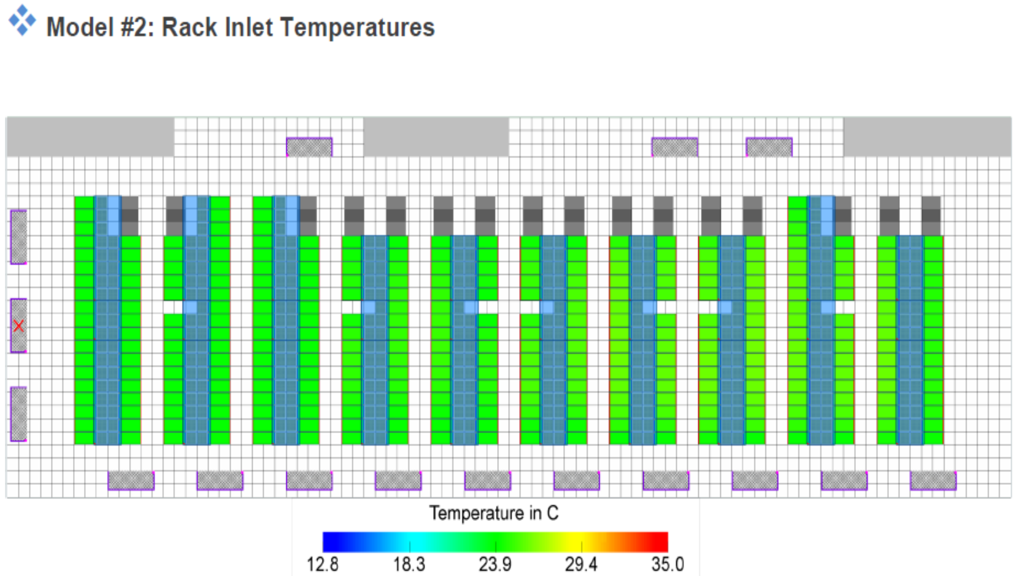


Fig. 20 shows data hall's floor layout after the Cold Aisle containment installation

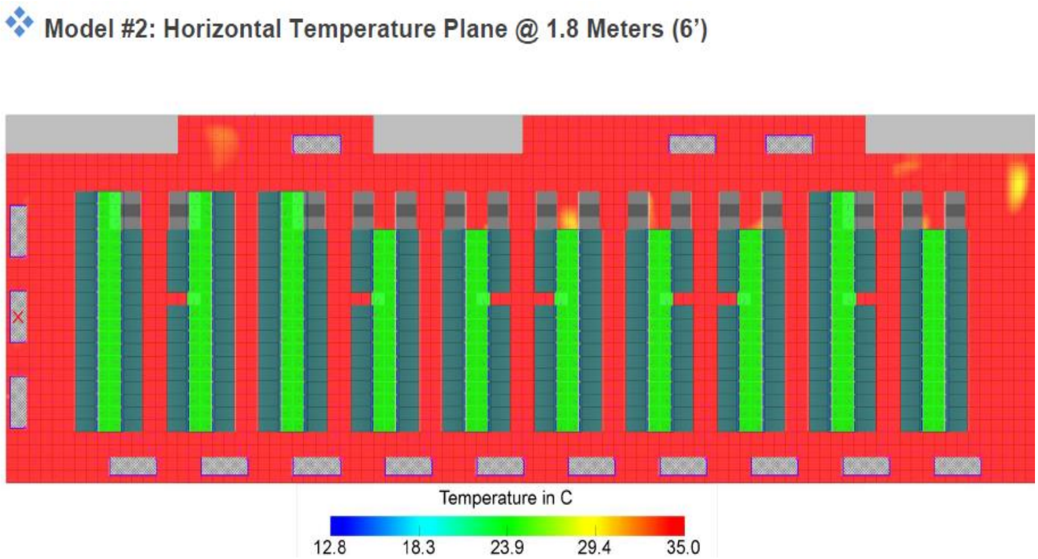


Fig. 21 shows data hall's floor layout after the Cold Aisle containment installation

Table 10 below shows Model 2 with containment

Total Airflow Demand	435,393 CMH
Total Airflow Supply	480,750 CMH
Total Estimated Heat load	2,438 kW
Range of Max Inlet Temperatures	
Above 27°C	0 racks
Between 24 °C and 27 °C	325 racks
Between 21 °C and 24 °C	0

Measurements that have been conducted in the data center.

The different type of measurements was taken from the building management system for this case study. Fig. 23a shows the trend (25th September – 05th October) of the measured return and supply temperatures for the CRAH unit 605, which has similar supply and return temperatures as for the rest of the CRAH units. Fig. 23b presents the measured return and supply average temperatures for the racks in the row highlighted in Fig. 22. The measurements were taken on 4th October 2021 from Viasala humidity and temperature meters. The observed average rack inlet temperature was 24.4 and rack return temperature was 35.3 degrees Celsius (the average values from all cold and hot aisle temperature sensors) (Fig. 22). These values were captured through Building Management System. The temperatures at the rack intake were within the ASHRAE-recommended allowable range. Unshown data in Fig. 22 and 23b indicates that temperature increases from bottom to top were higher than 1 °C, indicating if inefficient separation of hot and cold air or insufficient amount of air for the IT equipment.

4.3.3 Performance metrics

SHI and RTI performance measures were calculated for this case study. The average SHI values were in the range of 0.2-0.35 which is less than 0.4. This shows that there was little recirculation of hot air and that the streams of hot and cold air were properly separated.

	T_{return}	T_{supply}
CRAH603	32,9	22
CRAH604	35,5	22,1
CRAH605	34,5	22
CRAH606	34,6	21,9
CRAH607	35,3	22,1
Average	34,56	22,02

Table 11 shows CRAH units 603-607 return and supply temperatures

The average temperature difference at the rack level was calculated as 10.8 °C ($\Delta T\ IT\ equip$). In order to calculate the average temperature difference for the CRAH units, the supply and return temperatures were taken from the Building Management System for the CRAH units 603-607 (Table 11). The RTI was calculated as 116% which indicates the air recirculation issue. If the Cold aisle containment was properly sealed, without any open gaps between hot and cold air containment, the CRAH units’ airflow could be increased by 16% (=1-116%), however in our case, the high RTI is because of the unproperly sealed cold aisle containment which causes the hot air recirculation and mixing with the supplied cold air. Considering the performance metrics in this case study, the conclusion can be made that the Cold aisle containment requires significant improvements. Those will improve the overall PUE of the data hall, as the cold-water consumption will be reduced which will consequently decrease the usage demand of the chillers.

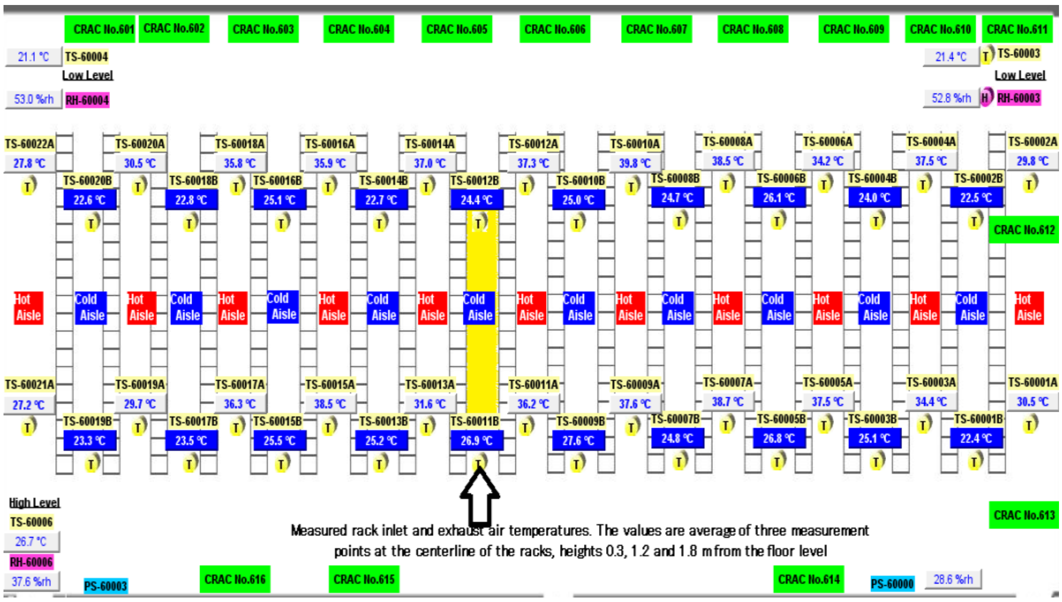


Fig. 22 shows data hall’s floor layout

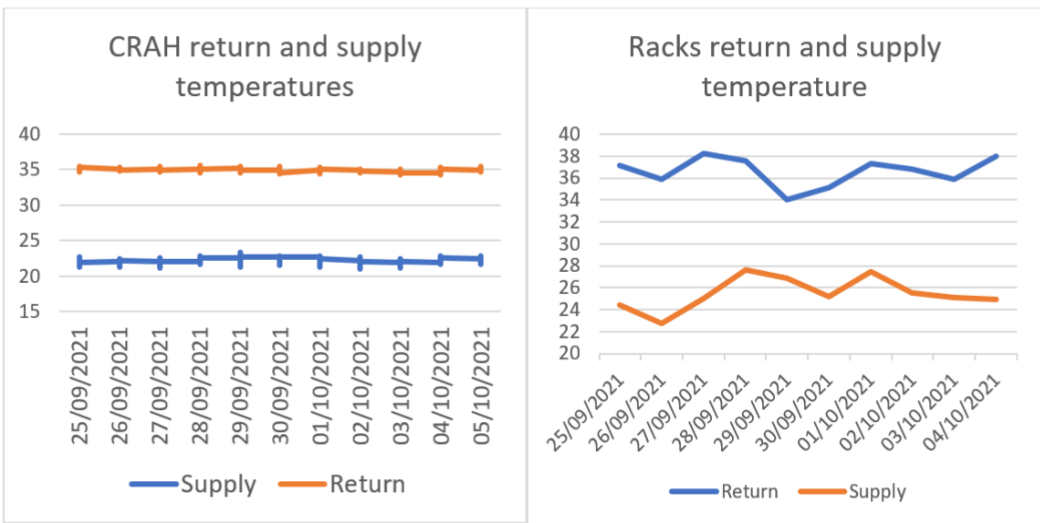
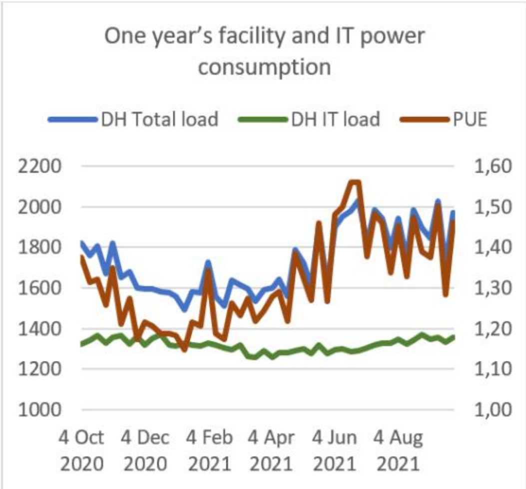


Fig. 23a shows CRAH return and supply temperatures and 23b - racks return and supply temperatures

4.3.4 Energy efficiency and usage breakdown

From the one year's facility and IT power consumption graphic (Fig. 24a), we can see that the PUE value was slightly above 1.23 in wintertime and around 1.44 during the summertime. The discrepancy was certainly caused by the free cooling mode, which was triggered usually when the outdoor temperature fell below 10-15 degrees Celsius. The average IT load was at 1319 kW over the year.



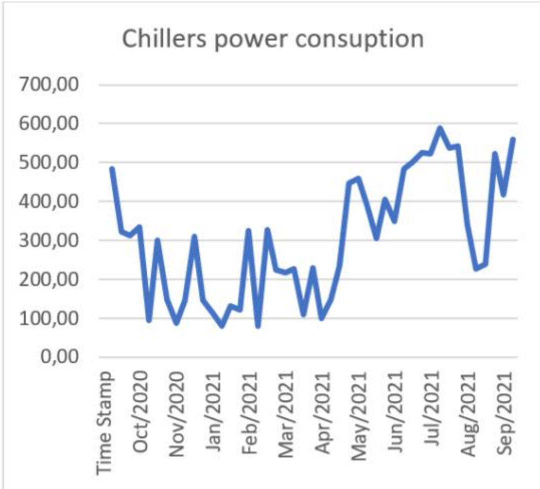


Fig. 24a shows PUE and DHs power consumption over the year and 24b (chillers power consumption over the year)

Power breakdown	kW
DH IT load	1319
CRAH units load	38,48
Chiller's load	304,66
Pumps	40,6
House services	20

Table 12 shows power consumption breakdown

4.4 Descriptive analyses

Since the power consumption of the Chillers may vary depending on the season (Fig 24b), the average value was captured from the electrical power management system. The calculated average PUE value is 1.31 which is below the max PUE advised by the German Digital Association. 22% of the total energy goes to the cooling system. Each year, about 97% of the total electrical energy is transformed to waste heat and subsequently rejected to the outside. The data center is not equipped with a heat recovery system. The massive quantity of heat rejected by the data center may be utilized for other reasons, such as heating the remainder of the building, providing hot water, or even being sold to energy markets. The quantity of heat that may be recovered from a data center is equal to the sum of the data center's extracted heat and the electrical power required to run the chiller. When free cooling is used, no electrical power is required for the chiller, and only extracted heat is available for re-use. The data center's waste heat can be captured using an additional condenser or heat exchanger. The recovered heat is typically in the form of 40.6–43.3°C hot water in ordinary buildings (Carrier Corporation, 2008). These temperatures are sufficient for the majority of applications found in commercial buildings. However, when free cooling is used, the created hot water from rejected heat may be significantly cooler than 40 degrees Celsius. This temperature of water might be utilized to warm up cold outdoor air before it enters the ventilation duct or to heat water for household usage. Nonetheless, approximately 14800 MWh per year energy may be utilized for the data center. In Germany, the typical dwelling consumes approximately 13.5 MWh/year for heating. This indicates that the reused heat can heat up to 1000 typical dwellings (=14800/13.5) on a yearly basis.

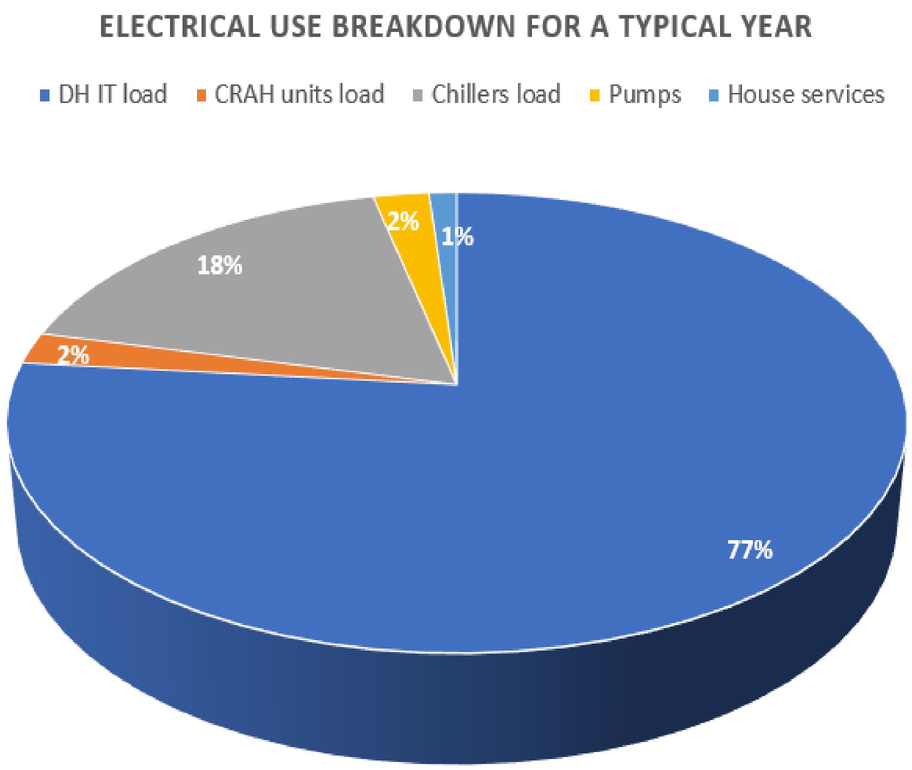


Fig. 25 shows Pie chart of electrical consumption over the one year.

4.5 Summary

This chapter is devoted to the optimization of cooling systems on the example of a real data center. In order to assess air management and cooling performance, a series of measurements was undertaken. The detailed CFD analysis was carried out in order to simulate the airflow properties of the data hall before and after Cold Aisle Containment installation. The simulation model shows the significant effectiveness of the cold aisle containment in terms of the prevention of hot air mixing with cold supply air. Energy balance was used to verify the accuracy of readings. The results indicate the data hall’s temperature was in within the ASHRAE allowed levels. PUE levels were between 1.2 and 1.5, depending on the mode of operation. There was no evidence of significant bypass air, although there was some recirculation air in the data hall. The cold aisle containment has many gaps that needs to be closed to prevent the air recirculation. There was no waste heat recovery system. Additional research is required to conduct a study of waste heat recovery systems and to compare data centers with and without waste heat recovery systems.

Performance metrics provide a significant potential for the whole data center business. They have the potential to provide a standardized method for defining and reporting diverse cooling systems. Two metrics are shown in this chapter: The Rack cooling index is a metric that indicates how successfully the system cools the devices in accordance with the manufacturer's standards. The Return Temperature index is used to determine the air management system's energy efficiency. When combined, they enable objective evaluation of the air management system's performance after extensive CFD modelling.

5. Waste heat recovery – case study

In the city called Odense in Denmark, the typical winter temperatures are around -1℃ and many of us have been unintentionally warming homes in this city every time when we click the "Like" button. Facebook operates the 2 buildings, 50,000 m2 data center in Odense, and the firm has partnered with local district heating provider Fjernvarme Fyn (FVF) to disperse heat created by the data center's IT equipment directly to the radiators in the surrounding town. The Data Centre is provided with a waste heat recovery (WHR) system which receives WHR water at 15 °C from a local district heating company FVF's Waste Heat Recovery plant to remove the heat extracted from the data hall Indirect Adiabatic Cooling IAC units and return the heated water back to WHR plant where its heat is extracted by heat pumps. The system is designed to provide the WHR plant with a return water temperature of up to 27°C where the temperature is raised further and delivered to the district heating network (Meta, 2020). The system circulates the WHR water through the WHR coils of data hall IAC units which have been selected for heat recovery and transfer the recovered heat back to the nearby WHR plant, approximately 1 km from the Data Center. There, its heat is extracted by heat pumps, with the resultant WHR chilled water generated being returned to the data halls' IAC units to be used for cooling.

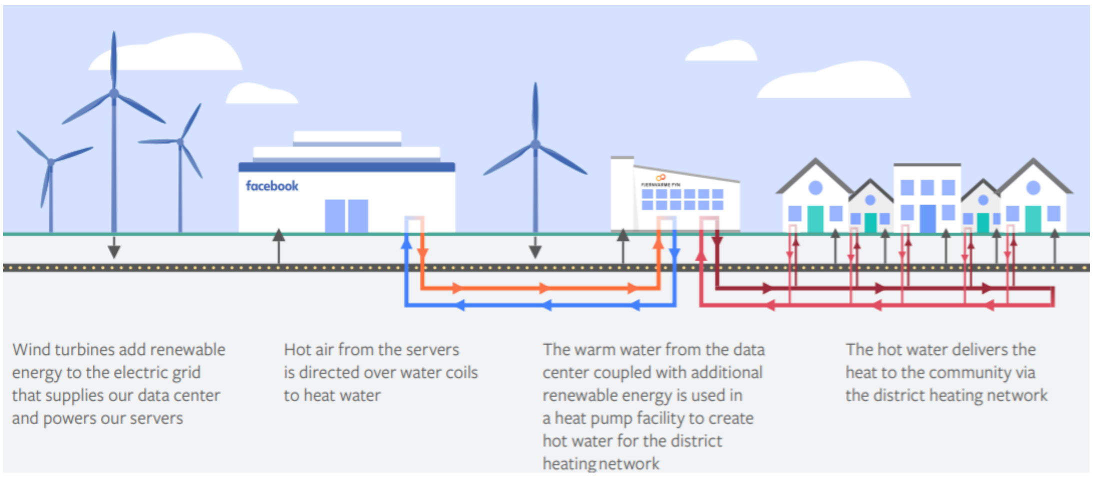


Fig. 26 shows Waste heat recovery system (Meta, 2020)

5.1 System description

The IAC units recirculate the air from the data hall through the IAC unit and the air will be cooled by the ambient air via air-to-air heat exchanger. The IAC units can operate in different modes:

- Outside Air (OA) mode, dry: Air to air cooling via the heat exchanger

- OA mode, adiabatic: Air to air cooling via heat exchanger plus evaporative cooling

- WHR mode: Air cooling via chilled water coil

Each IAC unit is equipped with 2 cooling coils for WHR with a combined corresponding cooling capacity of 392 kW.

Each data hall has its own distribution WHR pipeline connected to the IAC units. At the end of each data hall WHR pipeline is installed:

- A temperature sensor that registers the supply temperature

- An “end-of-line bypass valve” which is both valve motor controlled, and pressure differential controlled. The maximum flow through each valve is approx. 30 m^3/h.

- A differential pressure sensor that registers the current differential pressure at the end of the line.

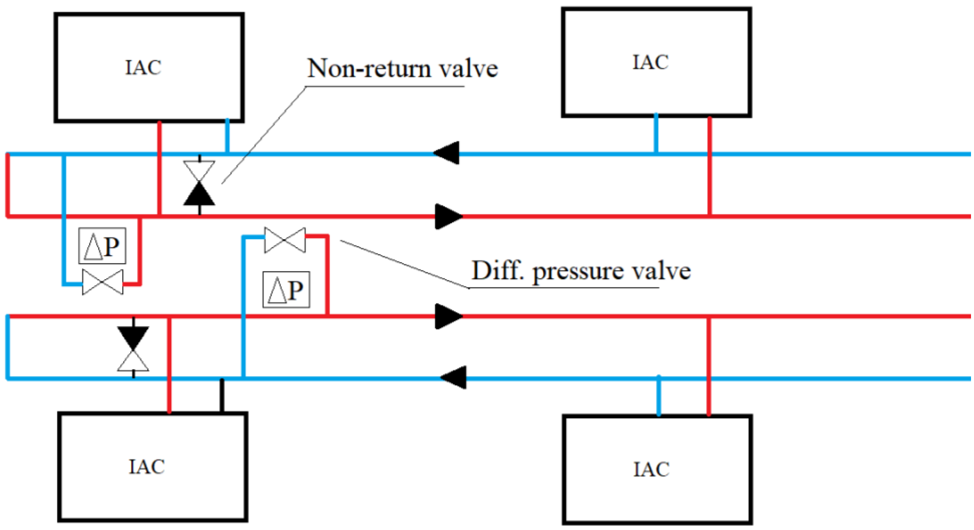


Fig. 27 shows WHR distribution piping end details with the illustration of the location of non-return valves at the end of the line.

Transient water hammer analysis for the WHR chilled water network has indicated that in full load situations a pump trip could cause pressure drops and on the roof of the data halls the pressure drops down to vapour pressure and pockets of vapour are formed. To avoid this risk, non-return valves are installed at the end of all lines. Fig. illustrates the location of the non-return valves. The addition of non-return valves in parallel with the end of line bypass valves will make it possible to have flow from the return to supply when the pressure in the return is higher than the pressure in the supply (negative differential pressure).

Prerequisites

Design WHR water conditions

- WHR water inlet temperature: 15°C (secured by FVF)
- WHR water return temperature: 27 °C (secured by Data Center)
- Peak WHR cooling demand per IAC unit: 392 kW
- Contractual supply of WHR capacity: 15 MW (from FVF)

Prerequisite for Transitioning IAC units to WHR mode

- The DC BMS is receiving *'WHR Plant Operational'* signal from the WHR Plant SCADA System.
- The supply temperature at the end of the WHR circuit line less than 17 °C for a period of 300 seconds.
- The differential pressure at the end of the WHR circuit line is ≥ 150 kPa.

When the WHR mode is not activated, all the systems end-of-line 2-port motorized flushing/pressure control valves (two per pair of data halls) are to be driven to their fully open positions. This is to provide an open circuit should FVF run and test their distribution pumps.

The Data Center BMS operator can, through the appropriate graphics pages at DC BMS, pre-select which IAC units are to be employed in WHR mode. However, pre-selection will not be allowed for any units connected to a WHR circuit whose end-of-line 2-port motorized flushing/pressure control valve is monitored to be closed as there will not be any flow through this circuit (A prerequisite to enable IAC units connected to the circuit to transpose to their WHR operating mode) and this could be the only circuit in which IAC units have been pre-selected.

WHR supply temperature to Data Halls Critically High alarm will be initiated after a delay time of 10 seconds if the supply temperature of the WHR circuit remains more than 4 °C above the required temperature (15 °C) for a sustained period of 300 seconds. The Data Center Operator can then stop the WHR mode and start-up the OA mode. The WHR-mode can be stopped by Data Center Operator as well by the FVF Operator. When the WHR system has been inactive for 168 hours, FVF will initiate a System Flush Mode Activated signal to the Data Center BMS for a set period of 15 minutes. While this signal is present, the WHR Plant SCADA System will start operating the WHR distribution pump set to ensure the value of the End of Line Minimum Pressure does not fall below the current set minimum value. While the system is not operating, the end of line Motorized Flushing/Pressure Control Valves are to be open, so the Data Center BMS takes no action upon receipt of this signal.

The distribution pipeline system between the WHR plant and the Data Center are divided into the above-ground pipeline and below the ground pipeline. The internal water volume of the pipeline for the Data Center buildings, the below-ground pipelines and the WHR plant:

- Above ground pipeline 71 m^3
- Below ground pipeline, Data Center site, first phase: 749 m^3
- Below ground pipeline, from Data Center site to WHR plant: approx. 227 m^3
- WHR plant, main piping incl. pressure vessel: approx. 74 m^3
- WHR plant, balancing tank: 75 m^3
- WHR plant, heat pump groups: approx. 12 m^3
- Total internal pipeline volume: 1467 m^3

When WHR-mode is stopped and after a long period with no circulation the internal water temperature for the above-ground placed pipeline can vary depending on the season (3 – 30 °C). For the below-ground pipeline, the temperature over time will equalize to the ground temperature (8-10 °C). At start-up of the circulation pumps, the water will return to the WHR plant like a train carriage with different temperatures. A strategy to avoid these various temperatures a constant basic circulation could equalize the temperature in the pipe system over time. When all end of line bypass valves are open a basic circulation of approx. 100-240 m^3/hr. would equalize the temperature over time. When the Data Center runs in OA mode, all the end line bypass valves are normally open. The WHR Plant control system operator can decide if the circulation shall be on constantly. If the WHR mode stand-still period will be long, then the circulation pumps could run e.g. one week.

The WHR plant will in the have heat pumps installed with a WHR chilled water capacity of 19-20 MW. The heat pumps are divided into 3 groups of each 3 heat pumps.

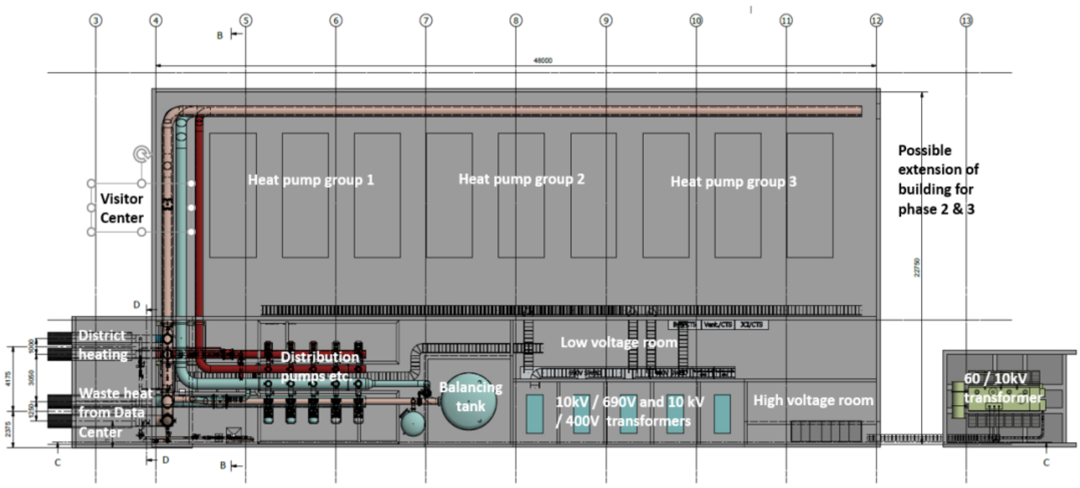


Fig. 28 shows Plan view of the WHR Plant with heat pumps, misc. technical rooms and piping for WHR and district heating

Conditions for start-up of Heat Pump groups:

The following conditions are a prerequisite for the heat pumps to start-up and subsequently be in operation without any failure:

- The following signals must be present for start-up: 'WHR Available' and 'IAC / WHR Mode Activated'.
- The water returned to the heat pumps from the distribution system or balancing (buffer) tank should be between 17 and 27 °C. A stable temperature must be ensured.
- The return temperature from the district heating network must not exceed 45 °C during all operating scenarios. It is important to be aware of too high inlet temperature to the heat pump by usage of the circulation connection.
- The differential pressure between district heating water supply and return must be stable at 0-0.5 bar. The differential pressure shall be 0 when performance and guarantee tests are carried out.

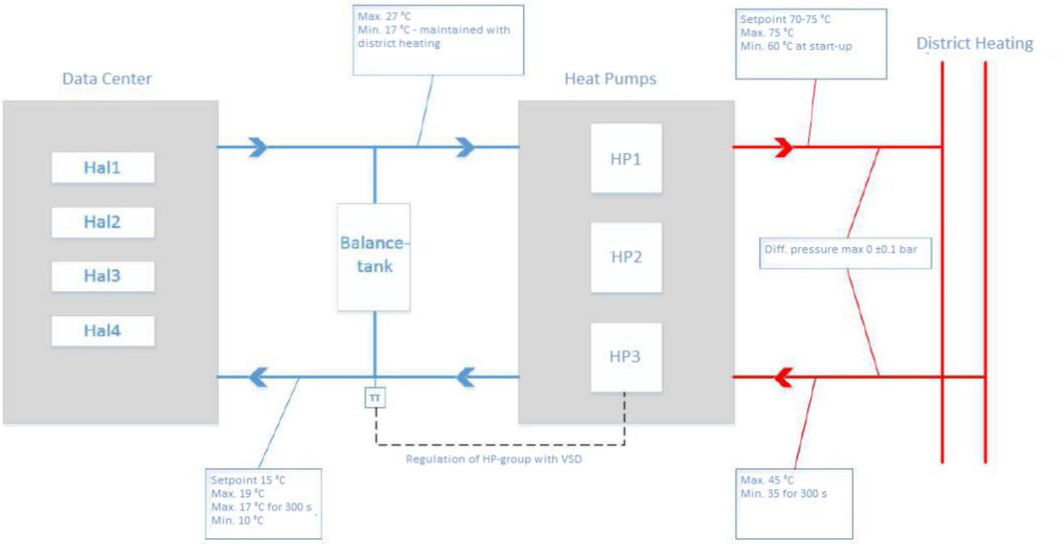


Fig. 29 shows Heat Pump start-up conditions

The number of heat pumps to be started depends on the *'Heat Pump Capacity'* announced to the Data Center BMS by WHR plant SCADA system. Each heat pump group have a cooling capacity of approximately 7 MW.

- When capacity request announced 0- 7 MW: One heat pump group is released
- When capacity request announced 7 - 14 MW: Two heat pump groups are released
- When capacity request announced 14 - 20 MW: Three heat pump groups are released

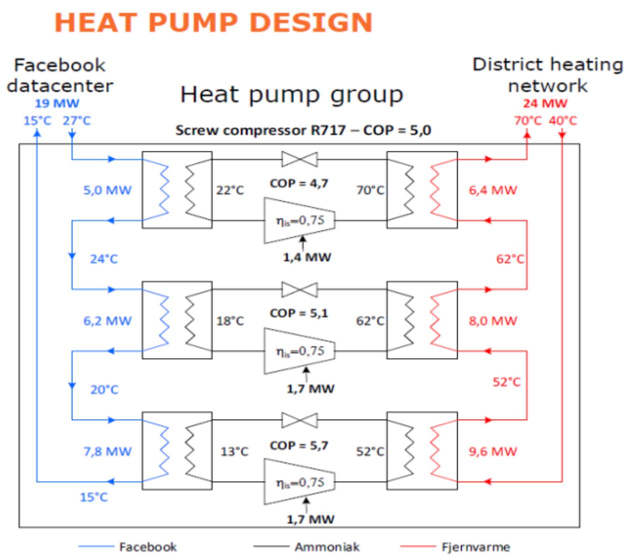


Fig. 30 shows typical heat pump design

This rule avoids unnecessary start-up of heat pump groups, which will shortly afterwards stop when the system has stabilized. To achieve the best coefficient of performance (COP), the capacity signal corresponds to the heat pump groups released are close to 100% loaded and to have a load of 0-20% on the heat pump group which regulate the capacity of all groups. This due to the supply of waste energy from the Data Center versus energy removed by the heat pumps will cause a critical low temperature drop in the evaporator and balancing tank.

Capacity range (WHR) critical to the operation:

- Capacity equivalent to 1 heat pump in operation, 20% of 7 MW: 0-1.4 MW
- Capacity equivalent to 2 heat pumps in operation, 20% of 7 MW (+ 7MW): 7-8.4 MW
- Capacity equivalent to 3 heat pumps in operation, 20% of 7 MW (7 + 7 MW): 14-15.4 MW

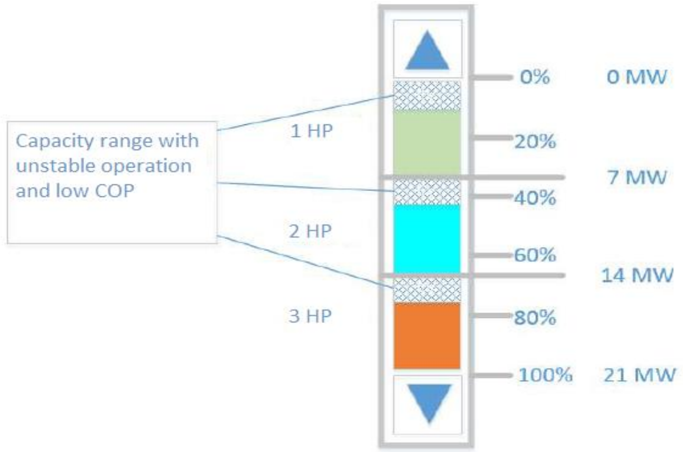


Fig. 31 shows HP Capacity range (WHR side) with unstable operation and low COP

Start-up sequence of Heat Pump groups

The starting sequence for the first heat pump group has the following steps:

1) The heat pump group is released for operation when the following conditions are met:

   a) The temperature in the return line from the balancing tank is min 17 °C and max 27 °C

2) Valves receive activation signal:

   a) The valves on the supply and return pipes of the evaporator opens.

b) On the condenser side (DH) there is a sequence of the bypass and the control valves:

i) While the HP builds up pressure, and thus temperature, the DH water is re-circulated via the bypass pipe until it reaches 60 °C

ii) Then the bypass valve closes

iii) The control valve to the DH opens and the flow enters the DH network supply side

3) Production pumps are released for operation:

a) WHR Plant chilled water pumps are released for operation with fixed speed at 80% (setpoint can be changed in the SCADA system)

b) The district heating pump is released and regulates the district heating water supply to 70-75 °C

4) Firstly, the high-pressure compressor starts with minimum capacity

5) 60 seconds later the intermediate pressure compressor starts with minimum capacity

6) When the temperature in the DH supply pipe reaches 60°C the by-pass valve starts closing while the control valve in the DH supply opens

7) When the by-pass valve is completely closed, the temperature in the DH return pipe is recorded. If it is less than 45 °C for 30 seconds, then the low-pressure compressor is turned on.

8) Now all the compressors of the group are in operation and regulate the flow until the WHR chilled water flow from the balancing tank is 15 °C.

When more HP groups are in operation:

9) 3 minutes later, the second heat pump group is released and has the same sequence as HP group 1.

10) 3 minutes later, the third heat pump group is released and has the same sequence as HP group 1.

The heat pumps regulate the capacity in accordance to the supply temperature to the Data Center at 15 °C. The supply temperature is measured immediately after the T-crossing at the balancing tank. The temperature and capacity regulation are controlled by the WHR Plant heat pump PLC. The 3 heat pumps in a group regulates up and down synchronously. Heat pump group no. 3 – which is equipped with variable frequency drives (VFD) - will start-up as the first after release signal and ramps up capacity to accommodate the supply temperature of 15 °C for the Data Center. If there still is a cooling need, the next heat pump group will start-up and regulate to 100% after which the heat pump group 3 will ramp down capacity to meet actual capacity need from Data Center and retain the supply temperature of 15 °C. If there still is a cooling need, the next heat pump group will start-up and regulate to 100% after which the heat pump group 3 will ramp down capacity to meet actual capacity need from Data Center and retain the supply temperature of 15 °C. The start-up process from cold (long stand still period) of the first Heat Pump group takes approximately 30 minutes before stable supply temperatures to the Data Center are achieved. The start-up of the following Heat Pump groups will take approximately 15 minutes.

When the WHR Plant wants to shut down or change the capacity request, it will send a signal to the Data Center informing the new capacity request or, if necessary, a total stop. The capacity signal always contains the 'Heat Pump Capacity' and 'Heat Pump Capacity Remaining from Target' signals. The sequence of shutting down the heat pump groups is the following:

1) Stop command or reduced capacity signal from FVF is initiated
2) Down-regulation of heat pumps:
   a) The heat pump group with frequency converter regulates compressors from current capacity gradually to 20%
      i) 60 seconds after: Low pressure compressor stops
      ii) 60 seconds after: Intermediate pressure compressor stops
      iii) 60 seconds after: High pressure compressor stops
   b) With fixed-speed heat pump groups, all compressors are regulated from 100% to 70%
      i) 60 seconds after: Low pressure compressor stops
      ii) 60 seconds after: Intermediate pressure compressor stops
      iii) 60 seconds after: High pressure compressor stops

3) Cold water pump on the heat pump group is stopped

4) Valves on cold water pipes are closed back and forth.

5) 10 minutes after: Hot water pump on the heat pump group stops

6) Bypass and district heating control valve is closed.

5.2 The WHR System – sequence of operation

The WHR Plant SCADA system manages and controls the WHR system. It controls the set of variable speed distribution pumps to meet the current demands of the system. Furthermore, it controls the heat pumps which deliver the waste heat to the FVF district heating network. However, the distribution pipework and all components within the boundaries of the Data Center campus are the responsibility of the Data Center facilities team. The WHR system is designed to recover up to approximately 20 MW of waste heat from the eight data halls. The minimum start-up heat required to be transferred from the WHR system to the FVF district heating network to prevent short cycling of the WHR Plant heat pumps is 1.4 MW corresponding to 4 IAC units in operation.

Information transactions between the Data Center (DC) BMS and WHR Plant SCADA systems are to be via hard-wired analogue and binary signals, with the interface point for the DC BMS being provided by a controller housed in the Security Lodge at the perimeter of the site. A multi-core cable is to be wired between the Security Lodge and the FVF heat pump building on the other side of the perimeter road. The continuous analogue signal in terms of kW provided by the WHR plant SCADA System to the DC BMS to calculate Heat Pump Capacity and identify the maximum amount of heat that can be accepted by the WHR Plant.

- Tempering the WHR water before start-up of the WHR mode
- WHR water return side

The heat pumps must have a WHR return water temperature of 17 to 27 °C. A pre-heating of the WHR water return by district heating supply water can ensure that the WHR water will be within these limits. This pre-heating system has a capacity corresponding to the capacity of one heat pump group. This will allow the heat pumps to be started before waste heat from the data center has arrived at the heat pumps. In addition, it can be used to balance out temperature fluctuations in the pipeline during startup.

The same connection from the DH supply line 70 °C will be used at the commissioning / SAT test to prove the capacity. This will be done without the WHR system distribution circuit being active (distribution pumps stopped). The DH supply line is sized to deliver a full capacity of one Heat Pump group (~7 MW).

If the WHR water supply temperature is below 15 °C the start-up of WHR mode can proceed immediately without starting the Heat Pumps. A feed forward temperature signal of the return WHR water can start-up the Heat Pumps in advance. The low limit for supply cooling temperature is 11 °C. If the WHR water supply temperature is lower than 11 °C, a heating process will start injecting DH supply water of 70 °C into the cooling supply piping before the distribution pumps. When the WHR water conditions are within the prerequisite values the WHR Plant control system operator can activate the '*WHR system operational'* signal.

WHR water temperature values:

WHR water Supply temperature: 11 °C ≤ TS ≤ 15 °C

WHR water Return temperature> 17 °C ≤ TR ≤ 27 °C

At start-up of the Heat Pumps the condenser water outgoing temperature to the DH supply piping can have a temperature ≤ 60 °C. In this situation the bypass regulation valve will be fully open. When the temperature reaches a temperature of 60 °C the bypass regulation valve starts to close, and the condenser water outgoing regulation valve will start open. The condenser water pump will run fixed 80% flow. When the outgoing temperature to the DH supply piping has reached 70 °C (Setpoint) the bypass valve changes position to fully closed, and the condenser water pump will start regulating the outgoing water temperature to the setpoint temperature of 70-75 °C. The outgoing regulation valve will be fully open.

The DC BMS operator can, through the appropriate graphics pages, pre-select which IAC units are to be employed in WHR mode. However, pre-selection will not be allowed for any units connected to a WHR sub-circuit whose end-of-line 2-port motorized flushing/pressure control valve is monitored to be closed as there will not be any flow through this circuit (a prerequisite to enable IAC units connected to the circuit to transpose to their WHR operating mode) and this could be the only sub-circuit in which IAC units have been pre-selected. When no IAC units are in

WHR mode the end-of-line 2-port motorized flushing/pressure control valve will be open.

A series of steps must be followed to activate the WHR Plant. The WHR Plant SCADA System will initiate the process by activating the *'WHR Available'* signal to the DC BMS (contact closed). The WHR Plant remains switched off. The DC BMS will accept the request by sending the *'IAC / WHR Mode Activated'* signal to the WHR Plant SCADA System. Before sending this signal, the DC BMS must ensure the end-of-line bypass valve is driven to the fully open position if the circuit contains any IAC units pre-selected for WHR. The end-of-line bypass valve must be driven to the fully closed position if the circuit does not contain any IAC units pre-selected for WHR. The WHR Plant SCADA System responds with a 'IAC / WHR Mode Activated Handshake' signal. The DC BMS shall not take any action when receiving this signal. Upon sending this signal, the WHR Plant will commence its start-up sequence. The distribution pumps will be speed regulated to maintain 150 kPa differential pressure at the critical consumer. The WHR Plant SCADA System will send the *'WHR Plant Operational'* signal when the heat pumps are in operation and when the supply temperature has been below 17 °C for at least 300 seconds (measured at the WHR Plant). Due to latency in the pipe system, the temperature at the end-of-line measuring points could be higher. For example, at 6 MW capacity (WHR chilled water) it could take 30 minutes before the end-of line temperature is 15 °C. The DC BMS can start transitioning the first four IAC units into WHR mode once the *'WHR Plant Operational'* signal has been received and the end-of-line conditions for temperature and differential pressure are fulfilled. The end-of-line bypass valve shall be driven to its fully closed position when the first IAC unit on the same circuit starts transitioning to WHR mode. The DC BMS shall send *the 'WHR Mode Activated'* signal to the WHR Plant SCADA System when the first IAC unit has started transitioning to WHR mode.

When the WHR system is fully operational, the DC BMS operator is able to manually take IAC units in and out of WHR operating mode via the appropriate graphics pages.

The decisions to bring any one or more IAC units into or out of their WHR operating mode should be based on the value of the energy freeboard that FVF have available, as given by the value of their '*Heat Pump Capacity Remaining from Target'*. If this value falls to less than the DC BMS must issue a level 2 alarm to the operator to reduce the number of IAC units currently in WHR operating mode.

The DC BMS must also issue a level 2 alarm to the operator to reduce the number of IAC units currently in WHR operating mode if the supply temperature of the WHR circuit remains more than 2 °C above the required temperature (15 °C) for a sustained period of 300 seconds. Energy meters are provided where each WHR sub-circuit enters each data hall. As well as displaying these values on the appropriate data hall graphics pages, they are all to be displayed on a single page together with the *'Heat Pump Capacity'* and the *'Heat Pump Capacity Remaining from Target'*. The *'Heat Pump Capacity Remaining from Target'* values are given by the WHR Plant SCADA System to the DC BMS to aid the operator in assessing which IAC units to bring into or take out of WHR operating mode. It should be noted that for each pair of data halls, there is an upstream energy meter measuring both halls in series with a downstream meter measuring the furthest of the two halls.

5.3 Summary

In this, last chapter, there was a detailed analysis and sequencies of operation of Facebook data center located Denmark where the waste heat from servers is used for warming homes in the city. In the study described in Chapter 4, there was no waste heat recovery process, which in turn leads to the fact that the energy that could provide 1000 apartments with heat during the year is released into the air from heat exchangers on cooling devices. The Facebook's Data Center provides every year about 100,000 MWh annually. The system absorbs heat generated by the data centre's hundreds of servers. Rather than capturing heat via air, the method channels it via copper coils filled with water. After that, the water is channeled to a heat pump facility managed by the local heating firm Fjernvarme Fyn, where it is distributed to households. Facebook predicts that the method will enable them to contribute 100,000 MWh of heat that would have been squandered otherwise to local communities.

6. Summary and Outlook

As our need for data keeps growing, the data center sector is exploding. These hubs serve as the foundation for a large portion of our economic, commercial, and social life. Additionally, they are some of the world's largest energy consumers. As governments commit to net zero carbon, environmental legislations are rapidly expanding; at the same time, Environmental, Social and Governance factors are advancing the business agenda. The sector is reacting, with several worldwide cloud and data center operators, but also smaller regional and national providers, pledging to climate neutrality by 2030. In our calculations, we found that most of the electricity consumed by the data center falls into the share of cooling systems. In this thesis, we pursue an energy-efficiency oriented treatments of cooling solutions in data center. In addition to reducing the impact on the environment, reducing electricity consumption of the cooling equipment will also lead to the fact that the end product, in the form of providing cloud services, will become much cheaper and more accessible. immersion cooling may be the next technology to be used in data centers. Probably soon we will witness this transformation. Similarly, the technologies and performance measurements that we discussed in Chapter 4 can be successfully used in other industries where it is required to measure the efficiency of energy consumption. The case study that we reviewed in Chapter 5 of how heat from a data center could be used to warm hundreds of homes. Equal to a couple of years ago, this energy was simply released into the air through heat exchange systems. Since data centers consume a lot of energy, it may be necessary to establish an energy efficiency standard at the state level, which will help achieve the goal of being environmentally neutral faster and more efficiently.

The study elaborated by several authors showed the many opportunities in terms of improving, standardizing, and recuperating energy and optimizing energy efficiency in data center cooling systems. The boundaries of this thesis have been defined by two main targets: increasing the efficiency of cooling resource provisioning through the use of dynamic control policies and enabling higher quality heat harvests that are critical for increasing the return on investment and sustainability indexes for the upcoming wave of data center heat recovery applications. It can be concluded that the use of a variety of energy-efficient and best-in-class hardware technologies and best practice topologies for cooling systems leads to further greening data center cooling.

Finally, given the increased emphasis on renewable energy and heat recovery in future data centers, the experimental research of the connection between data center infrastructures and possible heat recovery and storage applications is of special relevance.

7. References

3M. (2008). *3M™ Novec™ 7500 Engineered Fluid.* Product Information, Electronics Markets Materials Division.

3M. (2019). *Immersion cooling for data centers.* Electronics Materials Solutions Division. From 3M Science. Applied to Life: https://multimedia.3m.com/mws/media/1798606O/3m-immersion-cooling-brochure.pdf

3M. (2020). *Data Center Liquid Cooling Applications with 3M™ Novec™ Engineered Fluids.* Electronics Materials Solutions Division. From https://multimedia.3m.com/mws/media/1801351O/data-center-immersion-cooling-aplications-with-3m-novec-engineered-fluids-line-card.pdf

Armenta, F., Tchernykh, A., Cortés-Mendoza, J., Yahyapour, R., Drozdov, A., Bouvry, P., . . . Avetisyan, A. (2015). Heterogeneous Job Consolidation for Power Aware Scheduling with Quality of Service. *RuSCDays'15 - The Russian Supercomputing Days.* Moscow.

ASHRAE. (2021). *Equipment Thermal Guidelines for Data Processing Environments.* ASHRAE.

ASHRAE Journal. (1972). *ASHRAE Guide and Data Book.*

Balodis, R., & Opmane, I. (2017). *History of Data Centre Development.* IMCS - Institute of Mathematics and Computer Science.

Bash, C. E., Patel, C. D., & Sharma, R. K. (2011). Efficient Thermal Management of Data Centers—Immediate and Long-Term Research Needs. *HVAC and R Research*, 137-152.

Bash, C., Patel, C., & Sharma, R. (2006). Dynamic thermal management of air cooled data centers. *Thermal and Thermomechanical Proceedings 10th Intersociety Conference on Phenomena in Electronics Systems.* IEEE. doi:10.1109/ITHERM.2006.1645377

Bitkom. (2011). *Wie messe ich den PUE richtig?*

Bitkom e.V. (2021). *Klimaschutz und Energieeffizienz durch digitale Gebäudetechnologien.* Berlin.

Borderstep Institut. (2020). *Energiebedarf der Rechenzentren steigt trotz Corona weiter an.* Borderstep Institut. Dr. Ralph Hintemann. From https://www.borderstep.de/wp-content/uploads/2021/03/Borderstep_Rechenzentren2020_20210301_final.pdf

Carrier Corporation. (2008). *Heat recovery from chilled water system.* From https://www.shareddocs.com/hvac/docs/1001/Public/0A/04-581025-01.pdf

CERN. (2022). *Computing at CERN*. From CERN Accelerating science: https://timeline.web.cern.ch/timeline-header/109

Chainer, T., Schultz, M., Parida, P., & Gaynes, M. (2017). Improving Data Center Energy Efficiency With Advanced Thermal Management. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 1228-1239. doi:10.1109/TCPMT.2017.2661700

Chao, L., Guangming, C., Shuiquan, Y., & Yueqin, L. (2017). *Enclosed aisle effect on cooling efficiency in small scale data center.* Jinan: Elsevier. Retrieved 2022 from https://www.researchgate.net/publication/320758796_Enclosed_aisle_effect_on_cooling_efficiency_in_small_scale_data_center

Coles, H., & Herrlin, M. (2016). *Immersion Cooling of Electronics in DoD Installations.* Technical Report, Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Colocation America. (2014, November 5). *Is Hot or Cold Aisle Containment Right For Your Data Center?* From Data Center News: https://www.colocationamerica.com/blog/hot-vs-cold-aisle-containment

Computer History Museum. (2022). *ENIAC*. From Birth of the Computer: https://www.computerhistory.org/revolution/birth-of-the-computer/4/78

Daraghmeh, H., & Wang, C.-C. (2016). A review of current status of free cooling in datacenters. *Applied Thermal Engineering*. From https://www.sciencedirect.com/science/article/pii/S135943111632405X#!

DATA AIRE. (2022). *Precision Air Conditioning vs Comfort Cooling*. From https://www.dataaire.com/precision-vs-comfort-cooling/

Datacenter at CERN. (2016, Jam 28). *CERN Document Server*. (S. E. Bennett, Editor) From CERN Data Centre: http://cds.cern.ch/record/2127440/files/

Day, T., Lin, P., & Bunger, R. (2019). *Liquid Cooling Technologies for Data Centers and Edge Applications.* Whitepaper, Schneider Electric, Data Center Science Center.

Google Cloud. (2022, February 2). *System Architecture*. From Cloud TPU: https://cloud.google.com/tpu/docs/system-architecture-tpu-vm

Gough, C., Steiner, I., & Saunders, W. (2015). *Energy Efficient Servers.*

Hackenberg, D. (n.d.). *The Plenum Concept: Improving Scalability, Security, and Efficiency for Data Centers.* Technische Universität Dresden, Center for Information Services and High Performance Computing.

Heinemeier, K. (2017). Free Cooling: At What Cost? *ACEEE Summer Study on Energy Efficiency in Buildings*. From https://www.aceee.org/files/proceedings/2014/data/papers/3-1007.pdf

Herrlin, M. K. (2008). *Airflow and Cooling Performance of Data Centers: Two Performance metrics.* From http://ancis.us/images/SL-08-018_Final.pdf

Herrlin, M. K. (2010). *Data Center Air Management Research.* Pacific Gas and Electric Company. From https://www.etcc-ca.com/sites/default/files/OLD/images/am_rd_report_final_2010-09-22.pdf

IBM's ASCC introduction. (2022). From IBM Corporate Archives: https://www.ibm.com/ibm/history/exhibits/markI/markI_intro.html

Informa USA. (2009, September 28). *Dublin Data Center: Free Cooling Diagram*. From Data Center Knowledge: https://www.datacenterknowledge.com/inside-microsofts-dublin-mega-data-center/dublin-data-center-free-cooling-diagram

Informa USA. (2013, July 12). *Inside Facebook's Lulea Data Center*. From Data Center Knowledge: https://www.datacenterknowledge.com/inside-facebooks-lulea-data-center

Jin, C., Bai, X., An, Y., Ni, J., & Shen, J. (2020). Case study regarding the thermal environment and energy efficiency of raised-floor and row-based cooling.

Building and Environment. From https://www.sciencedirect.com/science/article/pii/S0360132320304856

Karki, K., Novotny, S., Radmehr, A., & Patankar, S. (2011). *Use of passive, rear-door heat exchangers to cool low to moderate heat loads.* ASHRAE Transactions.

Lawrence, A. (2020, April 27). *Data center PUEs flat since 2013.* From Uptime Institute: https://journal.uptimeinstitute.com/data-center-pues-flat-since-2013/

Li, J., Zhou, G., Tian, T., & Li, X. (2021). A new cooling strategy for edge computing servers using compact looped heat pipe. *Applied Thermal Engineering*. From https://www.sciencedirect.com/science/article/pii/S1359431121000557

Loyolan Ventures. (2021). *Supermicro Liquid Cooling Rear Door Heat Exchanger Inside Top*. From ServeTheHome: https://www.servethehome.com/liquid-cooling-next-gen-servers-getting-hands-on-3-options-supermicro/supermicro-liquid-cooling-rear-door-heat-exchanger-fans-and-top/

Lu, H., Zhang, Z., & Yang, L. (2018). A review on airflow distribution and management in data center. *Energy and Buildings*. From https://www.sciencedirect.com/science/article/pii/S0378778818315214

Meta. (2020). *Denmark data center to warm local community.* Case Study. From https://sustainability.fb.com/wp-content/uploads/2020/12/FB_Denmark-Data-Center-to-Warm-Local-Community.pdf

Meta. (2020). *Denmark data center to warm local community.* Facebook. From https://sustainability.fb.com/wp-content/uploads/2020/12/FB_Denmark-Data-Center-to-Warm-Local-Community.pdf

Meta. (2020). *Sustainability Report*. From https://sustainability.fb.com/report-page/data-centers/

Nada, S., Elfeky, K. E., Attia, A., & Alshaer, W. G. (2015). *Thermal Management of Electronic Servers under Different Power Conditions.* From https://www.researchgate.net/publication/281558995_Thermal_Management_of_Electronic_Servers_under_Different_Power_Conditions

Nadjahi, C., Louahlia, H., & Lemasson, S. (2018). A review of thermal management and innovative cooling strategies for data center. *Sustainable computing*, 14-

28. From https://www.sciencedirect.com/science/article/pii/S2210537917304067#!

Netlib . (2021). *LAPACK — Linear Algebra PACKage*. From https://www.netlib.org/lapack/

Ramdas, S., Rajmane, P., Chauhan, T., Misrak, A., & Agonafer, D. (2019). mpact of Immersion Cooling on Thermo-Mechanical Properties of PCB's and Reliability of Electronic Packages. From https://www.researchgate.net/publication/336616611_Impact_of_Immersion_Cooling_on_Thermo-Mechanical_Properties_of_PCB's_and_Reliability_of_Electronic_Packages

Rasmussen, N. (2005). Air distribution architecture options for mission critical facilities. *Elektron*, 68-71. From https://www.ee.co.za/wp-content/uploads/legacy/68,70-71.pdf

Schiavon, S., Ho Lee, K., Bauman, F., & Webster, T. (2010). *Influence of raised floor on zone design cooling load in commercial buildings.* University of California at Berkeley, USA, Center for the Built Environment. Retrieved from https://escholarship.org/uc/item/7bf4g0k1#page-1

Sullivan, R., Strong, L., & Brill, K. (2007). *Reducing Bypass Airflow Is Essential for Eliminating Computer Room Hot Spots.* White Paper, Uptime Institute. From https://www.42u.com/pdf/WhitePaper-ReducingBypassAirflow.pdf

Tech Power Up. (2022). *AMD Ryzen Threadripper PRO 3995WX*. From Tech Power Up: https://www.techpowerup.com/cpu-specs/ryzen-threadripper-pro-3995wx.c2314

The Green Grid. (2012). *PUE: A COMPREHENSIVE EXAMINATION OF THE METRIC.* White Paper. From https://www.thegreengrid.org/en/resources/library-and-tools/237-PUE%3A-A-Comprehensive-Examination-of-the-Metric

The Grin Grid. (2022). *About Us*. From THe Green Grid: https://www.thegreengrid.org/en/about-us

Tuma, P. (2010). *Open Bath Immersion Cooling In Data Centers: A New Twist On An Old Idea.*

U. S. Army Photo. (2022). *Historic Computer Images*. From United States Army Research Laboratory: https://ftp.arl.army.mil/ftp/historic-computers/

United States Environmental Protection Agency. (2021). *Understanding Global Warming Potentials.* From https://www.epa.gov/ghgemissions/understanding-global-warming-potentials

Uptime Institute. (2009). *Data Center Site Infrastructure.*

UptimeInstitute. (2021). *Explaining the Uptime Institute's Tier Classification System (April 2021 Update).* Matt Stansberry.

Urieli, I. (n.d.). Specific Heat Capacities of Air - Engineering Thermodynamics. From https://www.ohio.edu/mechanical/thermo/property_tables/air/air_Cp_Cv.html

Zhang, X., Lu, J., & Qin, X. (2013). BFEPM: Best fit energy prediction modeling based on CPU utilization. *Networking, Architecture and Storage (NAS), 2013 IEEE Eighth International Conference.* doi:10.1109/NAS.2013.12