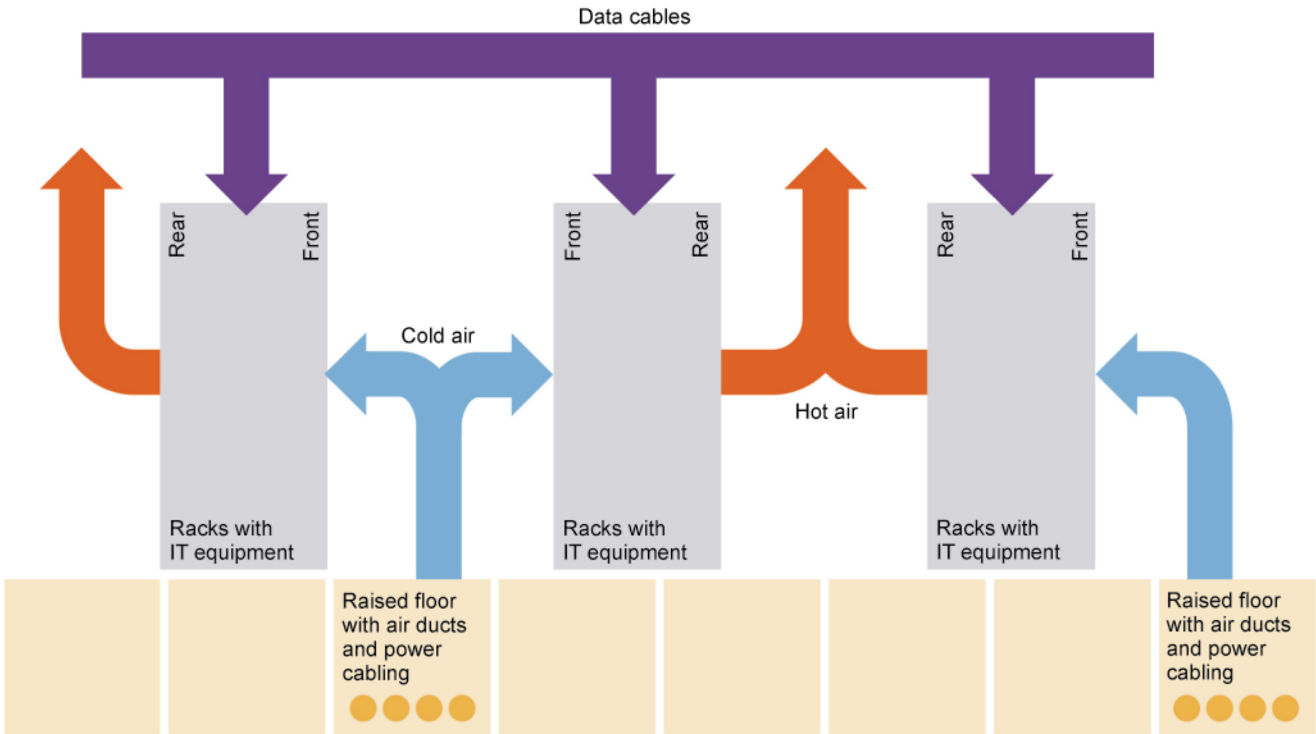
Green Sustainable Data Centres

Data Centre Facilities



This course is produced under the authority of e-Infranet: http://e-infranet.eu/

Course team
prof. dr. Colin Pattinson, Leeds Beckett University (United Kingdom),
course chairman and author of Chapter 1 and 7
prof. dr. Ilmars Slaidins, Riga Technical University (Latvia*),*
assessment material development: Study Guide
dr. Anda Counotte, Open Universiteit (The Netherlands),
distance learning material development, editor- in-chief
dr. Paulo Carreira, IST, Universidade de Lisboa(Portugal),
author of Chapter 8
Damian Dalton, MSc, University College Dublin (Ireland),
author of Chapter 5 and 6
Johan De Gelas, MSc, University College of West Flanders (Belgium),
author of Chapter 3 and 4
dr. César Gómez-Martin, CénitS - Supercomputing Center and
University of Extremadura (Spain),
author of Checklist Data Centre Audit
Joona Tolonen, MSc, Kajaani University of Applied Sciences (Finland),
author of Chapter 2

Program direction
prof. dr. Colin Pattinson, Leeds Beckett University (United Kingdom),
prof. dr. Ilmars Slaidins, Riga Technical University (Latvia)
dr. Anda Counotte, Open Universiteit (The Netherlands)

Hosting and Lay-out
http://portal.ou.nl/web/green-sustainable-
data-centres
Arnold van der Leer, MSc
Maria Wienbröker-Kampermann
Open Universiteit in the Netherlands

First edition 2014

Data Centre Facilities

Chapter 2

Data Centre Facilities

Joona Toolonen
Kajaani University of Applied Sciences

INTRODUCTION

This module *Data Centre Facilities* will focus on the aspects affecting the energy efficiency of a data centre. It aims to describe the basic mechanism of how energy is used in a facility and what actions will reduce the energy usage hence increase the efficiency and profitability of a facility.

This module is relevant to those administrating an IT facility who wish to increase the efficiency and decrease the costs of running such facility. The practical tasks within this module allow students to apply the theory to their own data centre.

LEARNING OBJECTIVES
After you studied this chapter we expect that you are able to
– describe the main building blocks of the facilities of a datacentre.
– describe the main energy consumption mechanism of a data centre
– understand the limitations of PUE, know other metrics and create an overall picture of different metrics working together
– understand how to make the most of energy measurement in a data centre
– formulate basic advice and plan how to reduce energy consumption in a data centre.

Study hints
The purpose of this chapter is to give an overview of the energy efficiency of facilities of a data centre. The workload is approximately 8 hours.

CORE OF STUDY

1 Facilities of a Data Centre

The number of data centres is increasing dramatically as modern cloud and web services gain popularity. At the same time energy price increases create pressure for service providers to offer more with less. Data centres need to be more efficient and all aspects need to be taken in consideration – including data centre facilities. [1]

1.1 CONFIGURATION OF A DATA CENTRE

The key purpose of a data centre is computing. Because the IT equipment needs a controlled environment, a data centre consists of a lot of other things than just IT equipment. Facilities need cooling, security, lighting and convenience equipment to run. All this equipment need electricity and are targets when reducing energy consumption in a data centre and aiming for greener environments. Some of the equipment takes relatively more energy than others so the possibility to save energy is greater in those areas. One of the biggest ones is cooling.

Cooling

A data centre *cooling* system removes the heat generated by the equipment. To remove heat, the cooling system must employ some hierarchy of loop systems, each bringing in a cold medium that warms up via some form of heat exchange and is somehow cooled back again. [1]

An open loop system replaces the outgoing warm medium with a cool supply from the outside, so that each cycle through the loop uses new material. A closed-loop system recirculates the same medium again and again, transferring heat to an upper loop in a heat exchanger, and eventually the environment. All systems must contain a route to the outside environment for ultimate heat rejection. [1]

A typical data centre cooling system, its power and cooling flow are described in figure 1.

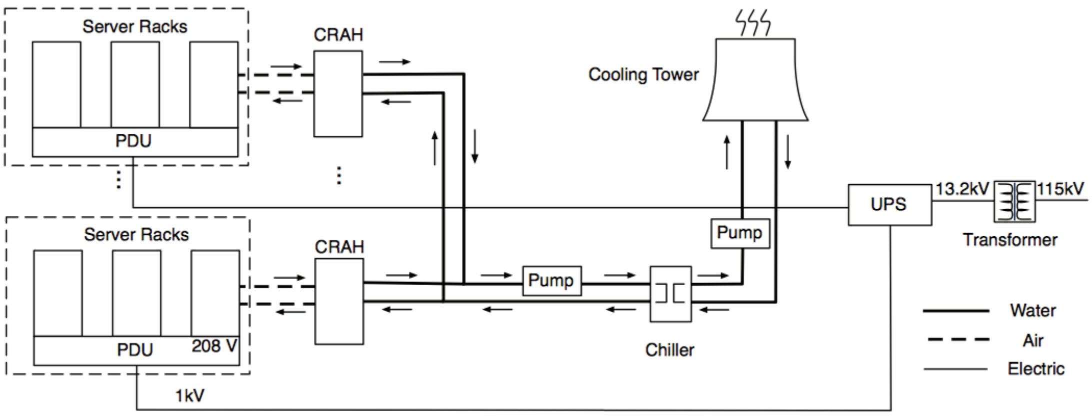


FIGURE 1 Power and Cooling flow in a datacentre (Pelley, et al., 2009)

Server Rack

Server Racks are standardized frames for mounting multiple equipment modules such as servers and switches. A power distribution unit (*PDU*) is a device fitted with multiple outputs designed to distribute electric power, especially to racks of computers and networking equipment located within data centres. [23] Servers and IT equipment are discussed more thoroughly in other Chapters.

Chiller

Computer room air handling unit (*CRAH*) is a device that uses circulating chilled water to remove heat and it must be used in conjunction with a *chiller*. A chiller is a device used to produce large volumes of chilled water that is then distributed to CRAH units. Usage of CRAH and chiller units leads to an increased consumption of water. [3] *Pumps* keep the fluids moving in the loop system and *a cooling tower* cools a water stream by evaporating a portion of it into the atmosphere. The temperature of the water drops significantly in the process.
In Figure 1 the cooling is provided by water which is cooled with air in a Cooling tower. Another possible cooling mechanism is air conditioning.

Uninterruptible power supply (UPS)

An *uninterruptible power supply (UPS)* is an electrical device that provides power to a data centre when the input source fails. UPS provides protection from input power interruptions practically instantaneously. UPS and power distribution in general is typically the second biggest source of power consumption in a data centre.

Transformer

A *transformer* takes care of transforming the input power source to a right level for data centre equipment.

1.2 COOLING

Because cooling consumes a substantial amount of energy, much research is on efficient cooling. Cooling is necessary because the IT equipment works in a specific temperature range and produces heat during operation: this heat may be sufficient to cause the overall temperature to exceed the safe operating range.
In figure 2 is described typical distribution of energy usage in a conventional data centre with a PUE of 2.0. [1] PUE is an acronym for Power Usage Effectiveness and is discussed more thoroughly *in section 1.2.2.*

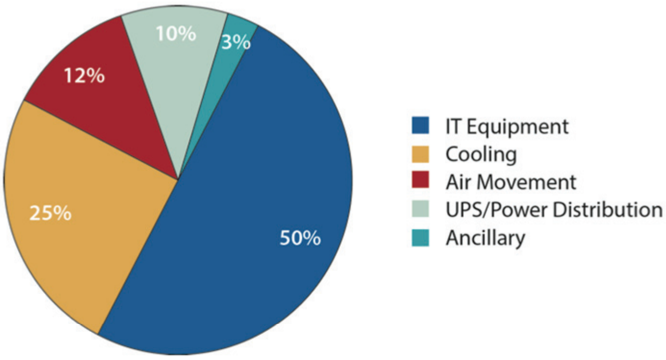


FIGURE 2 Power losses in a traditional (legacy) data centre. [1]

Figure 2 describes a traditional data centre whose PUE value is 2.0: If all cooling losses (25%) were eliminated, the PUE would drop to 1.26, whereas a zero-loss UPS system (10%) would only yield a PUE of 1.8. Typically, the worse a facility's PUE is, the higher the percentage of the total loss coming from the cooling system. Intuitively, there are few ways to introduce inefficiency into a power distribution system, but many more ways to do that for cooling. Much of this poor efficiency is caused by a historical lack of attention to efficiency, not by inherent limitations imposed by physics. Less than ten years ago, PUEs weren't formally used and a total overhead of 20% was considered unthinkably low. [1]

Recently, we have seen a focus on metrics for measuring Data Centre Energy Efficiency – also metrics beyond PUE. The emphasis is on cooling because of its importance when considering investments for saving energy in a data centre. This chapter introduces some of the best and most well known practices of biggest data centre providers and the latest research results into cooling data centres. General guidelines, tools and principles are presented for how to improve overall performance of a facility.

1.2.1 *Basics of Cooling*

99% of energy used to run IT equipment turns into heat

In almost every case, *99% of energy* used to run IT equipment in a data centre turns into *heat*, hence cooling plays an important role. Heat is a form of energy that can be measured relative to any reference temperature, body or environment. Temperature is a measurement of heat energy: different measures of heat intensity are *Celsius*, *Fahrenheit* and *Kelvin*. [3]

The second physical phenomenon related to cooling is *pressure*. Pressure is a basic physical property of *gas* and measured as the force exerted by the gas per unit area on surroundings.
The third physical variable related to cooling is *volume*, which means the amount of space taken up by matter. [3]

Properties of heat energy

There are three related *properties of heat energy*:
1 *Heat can only flow in one direction: from hot to cold.*
2 *Heat energy cannot be destroyed.*
3 *Heat can be transferred from one object to another object.* The three transfer methods are conduction, convection and radiation.

Conduction

In *conduction* heat energy is transferred due to temperature differences within a body or between bodies in thermal contact without the involvement of mass flow and mixing. It is the mode of heat transfer through solid barriers. Materials have a value called thermal conductivity (*W/mK*) and its value is higher for good electrical conductors and single crystals like diamond. Next in order are alloys of metals and non-metals. Liquids have lower conductivity than these materials. Gases, like air, have the lowest thermal conductivity values as seen in figure 3. [15]

Material	*Thermal conductivity, W/mK*
Copper	386.0
Aluminium	204.2
Glass	0.67
Water	0.60
Air	0.026

FIGURE 3 Thermal conductivity of some materials at 293K [15]

Convection most important

In *convection* heat transfers energy is transferred as heat to a flowing fluid or gas at the surface over which the flow occurs. This mode is basically conduction in a very thin fluid layer at the surface and then mixing caused by the flow. The heat flow is independent of the properties of the material of the surface and depends only on the fluid properties.
Convection is the *most important* mode of transferring heat from a surface.
Convection is not as pure a mode as conduction or radiation and hence involves several parameters. If the flow is caused by external means like fans, then the mode is known as *forced convection*. If the flow is due to the buoyant forces caused by temperature difference in the fluid body, then the mode is known as *free* or *natural convection*. In the design process thus the convection mode becomes the most important one in the point of view of application. [15]

Radiation

Thermal *radiation* is part of the electromagnetic spectrum in the limited wave length range of 0.1 to 10 μm and is emitted at all surfaces, irrespective of the temperature. No medium is required for radiative transfer of heat. [15]

Though each mode of heat transfer was discussed separately above, in practice all the three modes of heat transfer can occur simultaneously. Most of the time conduction and convection modes occur simultaneously when heat from a hot material is transferred to a cold fluid or gas through an intervening barrier. [15]

REFLECTION 1
What does the physics of heat mean in practice in a data centre?

Ideal Gas Law

In a data centre we deal with hot and cold air, gases that obey the *ideal gas law*. The relation between pressure (*P*), volume (*V*) and temperature (*T*) is known as the Ideal Gas Law, which generally states that

$$\frac{PV}{T} = \text{Constant}$$

This means that the increase in temperature results in a proportional increase in volume. If volume is constant, an increase in temperature results in a proportional increase in pressure. Inversely, if volume is decreased and pressure remains constant, the temperature must decrease. [3]

Removal of heat: CRAC, HVAC

Heat is traditionally *removed* from data centres via a refrigeration cycle mechanism. Computer room air conditioning unit (*CRAC* – some occasions called more generally heating, ventilation and air conditioning – *HVAC* unit) is a device that uses a self-contained refrigeration cycle to remove heat from the room and direct it away from the data centre. [3]

1.2.2 *The Nature of Humidity*

Humidity control

Static electricity

Humidity control is essential to high availability since proper humidity levels reduce *static electricity*. The movement of dry cooling air can be a source of static electricity. Electrically conductive and slightly wet air reduces the potential for electro-static discharge. [3]

Dew point

Relative humidity means the actual amount of water vapour in the air relative to the maximum amount of water vapour the air can hold at a given temperature. The d*ew point* is the temperature (varying according to pressure and humidity) below which water droplets begin to condense and dew can form. Relative humidity, dew point and temperature are all related, see figure 4. Therefore, to control the IT environment humidity and temperature one can either maintain the relative humidity, or maintain the dew point temperature at the CRAC level. [3]

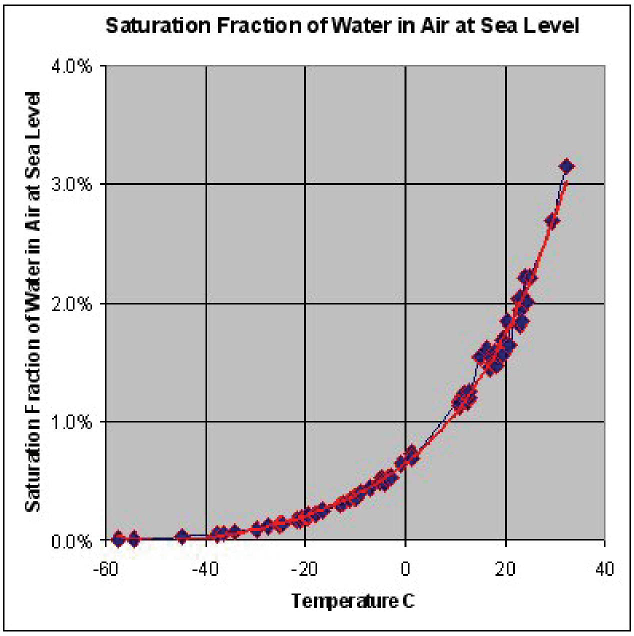


FIGURE 4 Dew point: This graph shows the maximum percentage, by mass, of water vapour that air at sea-level across a range of temperatures can contain. (Wikipedia, retrieved 030114)

If a data centre is controlled based on relative humidity, the increase in temperature causes more moisture to be added. In a data centre with two CRAC units with the same relative humidity settings (e.g. 45%), if the air in that room is returning to the CRACs at different temperatures, the higher temperature return air will have more water added to it by the humidifier in the CRAC unit than the lower temperature return air will. When a room contains several CRAC units set to maintain the same relative humidity setting the unequal addition of moisture among the units can eventually trigger one or more of the units to go into dehumidification mode. The other CRAC units will detect the resulting drop in humidity and will increase their own humidification to compensate. In an unmonitored room containing several CRAC units, it is possible to have half the room's cooling units adding humidity while the other half work to reduce it.

Demand fighting

This condition is known as *demand fighting*. [3]

REFLECTION 2

How is humidity controlled in your data centre and why is it important?

Uncoordinated CRACs fall short of cooling capacity and cause higher operating costs. CRAC units have four operating modes: *Cooling*, *heating*, *humidification* and *dehumidification*. While two of these operating modes may occur at the same time (e.g. cooling and dehumidification), all systems within a defined area should always be operating in the same mode. Demand fighting can have drastic effects on the efficiency of the CRAC system leading to a reduction in the cooling capacity, and is one of the primary causes of excessive energy consumption in IT environments. If not addressed, this problem can result in a 20 - 30% reduction in efficiency. [3]

Dew point control

Dew point control of IT environment is more cost effective than relative humidity control, as it greatly reduces the frequency of demand fighting. This is due to the fact that as air temperature in an IT environment increases its dew point stays the same. For example, air at 38°C exiting a piece of computer equipment has exactly the same dew point as the 27°C air entering the computer. Maximum cold aisle air temperature defined by the ASHRAE (American Society of Heating, Refrigerating and Air Conditioning Engineers) is 27°C.
Relative humidity and measured air temperature are always related to any specific dew point temperature. When several CRAC units are set to maintain humidity via dew point large differences in return air temperature will not drive excessive humidification or dehumidification in different units. All cooling units simply maintain humidity based on actual amount of water required in each pound of air that passes through the unit. [3]

People in a data centre, and leaking or un-insulated water pipes can increase humidity in the IT environment, while the air conditioning process and infiltration by drier outside air can decrease humidity. Minimizing the factors that affect humidity internal to the IT environment is equally as important as controlling external factors. By controlling both the internal and external factors that affect humidity levels in the data centre, the performance of the systems that have been designed to regulate humidity may be maximized. [3]

2 The Green Grid XUE Family of Metrics

Minimise the energy consumption

In Chapter 1 we introduced the performance metrics relevant to a data centre. We have seen that the IT equipment works in a controlled environment. The purpose is to *minimise the energy consumption* from as well the IT equipment as the auxiliaries. In this section we will explain what these metrics are. In this chapter we focus on the improvement of cooling and in Chapters 3 and 4 on the improvement of the IT equipment. In Chapter 5 and 6 we learn the best way to collect and control these metrics.

The Green Grid introduced PUE as a metric in 2007. [17] Over the years the limitations of PUE were recognized and in 2010 [6] The Green Grid introduced two other metrics CUE and WUE to improve the metrics of the energy efficiency of a data centre.

REFLECTION 3
How are CUE, PUE and WUE metrics related to each other?

2.1 POWER USAGE EFFECTIVENESS METRIC

Power Usage Effectiveness (*PUE*) reflects the quality of the data centre building infrastructure itself, and captures the ratio of total building power to IT power (the power consumed by the actual computing and network equipment, etc.).

$$\text{PUE} = \frac{\text{Facility power}}{\text{IT Equipment power}}$$

Relationships between data centre components and PUE are described in the following Figure 5.

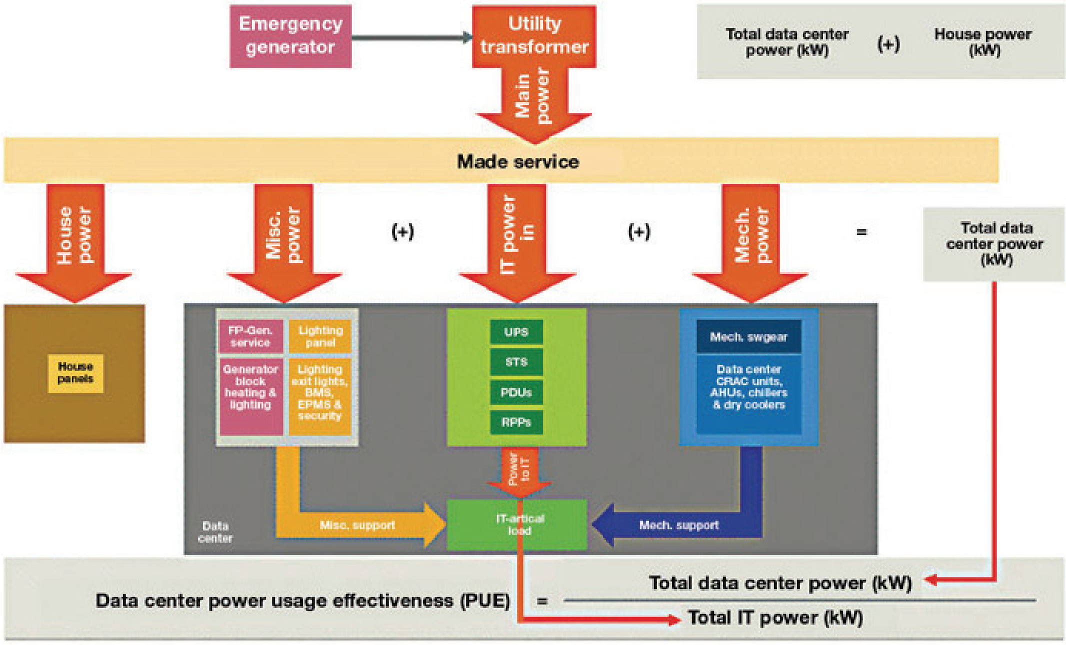


FIGURE 5 Data Centre Components and PUE (Salim, 2009)

There are four categories of PUE measurement, moving from lower accuracy in PUE Category 0, to higher accuracy in PUE Category 3. The categories differ in what is measured (instantaneous power vs. cumulative energy) and where it is measured, as shown in figure 6 [22]

	PUE Category 0*	PUE Category 1	PUE Category 2	PUE Category 3
IT energy measurement location	UPS output	UPS output	PDU output	IT equipment input
Definition of IT energy	Peak IT electric demand	IT annual energy	IT annual energy	IT annual energy
Definition of total energy	Peak total electric demand	Total annual energy	Total annual energy	Total annual energy

* For PUE Category 0, the measurements are electric demand

FIGURE 6 Different PUE categories [22]

The ideal PUE is 1. In that case the energy is consumed only by the IT equipment. The research aims to achieve the situation in which the auxiliaries use as little energy as possible. This leads to optimization of cooling techniques, for example hot and cold aisles: see section 4.1. But efficient data centres seek to optimize further. They aim to use as little energy as possible in the IT equipment. This is achieved by consolidation, virtualization and other techniques. This will be discussed in Chapter 4.

PUE has gained a lot of traction as a data centre efficiency metric since widespread reporting started around 2009. Historically, the PUE for the average data centre has been embarrassingly poor. According to a 2006 study [4], 85% of current data centres were estimated to have a PUE of greater than 3.0. In other words, the building's mechanical and electrical systems consumed twice as much power as the actual computing load. Only 5% had a PUE of 2.0 or better. [1]

Cold/hot aisle containment: CAC and HAC

A subsequent EPA survey of over 100 data centres reported an average PUE value of 1.91 and a 2012 Uptime Institute survey of over 1100 data centres covering a range of geographies and data centre sizes reported average PUE values between 1.8 and 1.89. The distribution of results is shown in Figure 7. The study noted *cold/hot aisle containment* (CAC/HAC) and increased cold aisle temperature as the most common improvements implemented. Large facilities reported the biggest improvements, and about half of small data centres (<500 servers) still were not measuring PUE. [1]

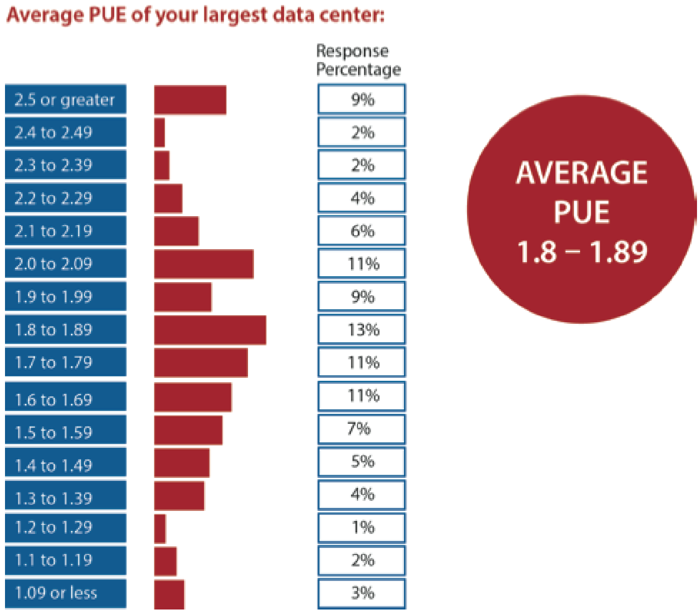


FIGURE 7 Uptime Institute survey of PUE for 1100+ data centres [1]

Measuring only PUE values has its faults and many values published by data centre operators are not directly comparable. Sometimes PUE values are used more in marketing documents to show best-case values than to describe the factual energy efficiency. [1]

The biggest factors (based on Barroso et. al. [1]) that can skew PUE values are:

– *Not all PUE measurements include the same overheads*. For example, some may include losses in the primary substation transformers, or losses in wires feeding racks from PDUs, whereas others may not.

– *Instantaneous PUEs differ from average PUEs*. Over the course of a day or a year, a facility's PUE can vary considerably. For example, during a cold day the PUE might be very low, but during the summer it might be considerably higher. Generally speaking, annual averages are most useful for comparisons.

– *Some PUEs aren't real-world measurements*. Often vendors publish 'design' PUEs that are computed based on optimal operating conditions and nominal performance values, or publish a value measured during a short load test under optimal conditions.
– *Some PUEs values have higher error bars* because they are based on infrequent manual readings, or on poorly placed meters that force some PUE terms to be estimated instead of measured.

REFLECTION 4
What are pros and cons of using a PUE value for a data centre?

Green Grid guidelines

For PUE to be a useful metric, data centre owners and operators should adhere to The *Green Grid guidelines* in measurements and reporting, and be transparent about the methods used in arriving at their results. Also all PUE values should be measured in real time. Not only does this provide a better picture of diurnal and seasonal variations, it also allows the operator to react to unusual readings during day-to-day operations. [1]

As said, The PUE value should not be the only metric for energy efficiency. It is not always indicating better energy performance, because for example PUEs typically worsen with decreasing load. Assume a particular data centre that runs at a PUE of 2.0 at 500 kW load versus a PUE of 1.5 at 1 MW load. If the same workload can be run with 500 kW of load (e.g. with newer servers), that clearly is more energy efficient despite the inferior PUE. Still the widespread adoption of PUE measurements has arguably been the driver of the biggest improvements in data centre efficiency in the past 20 years and is therefore justifiable metric to exist. [1]

2.2 CARBON USAGE EFFECTIVENESS

Carbon footprint

In Chapter 1 we discussed *carbon footprint*. When energy is obtained from fossil fuels, energy consumption and CO_2 production are of the same magnitude and can be expressed as carbon footprint. A low carbon footprint is the same as high Carbon Usage Effectiveness.
For data centres that obtain their entire power source from the energy grid and generate no local CO_2, Carbon Usage Effectiveness (*CUE*) is defined as follows:

$$\text{CUE} = \frac{\text{Total } CO_2 \text{ emissions caused by the total Data Center Energy}}{\text{IT Equipment Energy}}$$

The components for the loads in this and following metrics can be described as follows:
1 *IT Equipment Energy.* This includes the load associated with all of the IT equipment, including compute, storage, and network equipment, along with supplemental equipment such as KVM (keyboard, video and mouse) switches, monitors, and workstations/laptops used to monitor or otherwise control the data centre.

2 *Total Data Centre Energy.* This includes the IT equipment energy and everything that supports the IT equipment load, including:
– Power delivery components such as UPS, switch gear, generators, PDUs, batteries, and distribution losses external to the IT equipment.
– Cooling system components such as chillers, CRACs, direct expansion air handler (*DX*) units, pumps, and cooling towers.
– Other miscellaneous component loads such as data centre lighting.
– Total Data Centre Energy also includes other energy types beyond electricity, such as the natural gas that runs an absorption chiller.
3 *Total CO_2 Emissions.* This component includes the CO_2 emissions from local and energy grid–based energy sources. Ideally, the CO_2 emissions will be determined for the actual mix of energy delivered to the site (e.g., the electricity may have been generated from varying CO_2-intensive plants—coal or gas generate more CO_2 than hydro or wind. The mix also must include other energy sources such as natural gas, diesel fuel, etc.). The total CO_2 emissions value will include all Greenhouse gasses (GHGs), such as CO_2 and methane (CH_4). All emissions will need to be converted to 'CO_2 equivalents.' [6]

2.3 WATER USAGE EFFECTIVENESS

Water footprint

In Chapter 1 we discussed the importance of water and the *water footprint*.
The metric for water usage in the data centre is defined at a high level as:

$$\text{WUE} = \frac{\text{Annual Water Usage}}{\text{IT Equipment Energy}}$$

The units of WUE are liter/kilowatt-hour (L/kWh).

ASHRAE

5.5°C dew point.

Water use associated with the data centre is a complex topic at many levels. With WUE, the issue of a 'source-based' versus 'site-based' metric must be considered. The main issue is that water use or changes to a site's water use strategy generally affects other site use parameters and also can affect the supply chain for different utilities. A reduction in water use on-site can be accomplished in a number of ways. The most attractive way is simply to employ optimal design, then increase operational efficiencies and tune the existing systems. Re-commissioning a facility can accomplish this. The industry is replete with horror stories of data centres where one computer room air conditioning (CRAC) unit is dehumidifying while another is humidifying—together wasting both water and energy. In addition, many data centres have yet to take advantage of the *ASHRAE* 2008 extended environmental envelope where recommended *minimum humidity* levels have been reduced to *5.5°C dew point*. [5]

The use of tap water can achieve low-energy cooling even if no local sources are available. This has lead to some data centres becoming increasingly energy-efficient at the cost of wasting potable water. WUE has not yet achieved similar success as PUE. The situation is improving however, as based on Uptime's study, 34% of responders are already collecting water usage data. [2]

2.4 OTHER METRICS

Other metrics among the xUE family have been defined. The following section describes *Net Power Usage Effectiveness*, *corporate average Data Centre Efficiency* and *Computer Units per second* metrics. Adaptation and popularity of these different metrics compared to xUE varies.

2.4.1 *Net Power Usage Effectiveness*

Reclaimed energy

The Net Power Usage Effectiveness (*NPUE*) metric was introduced by Anders Greijer from Kungliga Tekniska Högskolan in 2010. NPUE tries to take *reclaimed energy* of a data centre into account. Following formula for calculating NPUE values has been presented:

$$\mathrm{NPUE} = \frac{E_{in} - E_{out}}{E_{it}}$$

Where E_{in} is energy input for a data centre, E_{out} energy output of a data centre and E_{it} energy consumption of IT equipment of a data centre.

PUE is not able to compare different kinds of cooling solutions and does not take into account the large amount of heat produced in a data centre. This limitation of the existing index gives rise to calls for an improved way of measuring energy efficiency in data centres.
Greijer suggests a different index (NPUE), which can compare different cooling solutions such as *district cooling* compared to chillers. The suggested index also gives a better number to facilities that re-use the heat energy produced by the servers and the chillers. [19]

The major changes from PUE to NPUE can be summed up as:
– NPUE measures the net energy flow to and from the data centre, where PUE measures the electric energy or power delivered to the data centre.
– NPUE measures the energy used over a period of 12 months.
The net energy to and from the data centre incorporates energy in other forms than electric energy, such as cooling in the form of district cooling to the data centre and heat energy delivered from the data centre. [19]

REFLECTION 5
What downsides and upsides do multiple measurement metrics and tools have?

2.4.2 *Corporate Average Data Centre Efficiency*

Corporate Average Data Centre Efficiency (*CADE*) metrics was introduced by McKinsey & Company in 2008. CADE defines five levels that aim to describe combined energy efficiency of IT and facilities. CADE is defined as follows:

$$\mathrm{CADE} = AU_{fac} \times EE_{fac} \times AU_{IT} \times EE_{IT}$$

where AU_{fac} is facility utilization, EE_{fac} is facility energy efficiency, AU_{IT} is IT utilization and EE_{IT} is IT energy efficiency.

To measure how effectively the data centre uses energy coming into the facility, CADE takes the amount of power consumed by IT, or the IT load, and divides it by the total power consumed by the data centre. To determine how fully the physical equipment installed at the facility level is being used, the CADE formula divides the IT load by the facility's total capacity. This Facility Efficiency measure is then multiplied by the average CPU server utilization and yields the organization's CADE rating. An example is shown in figure 8. [20]

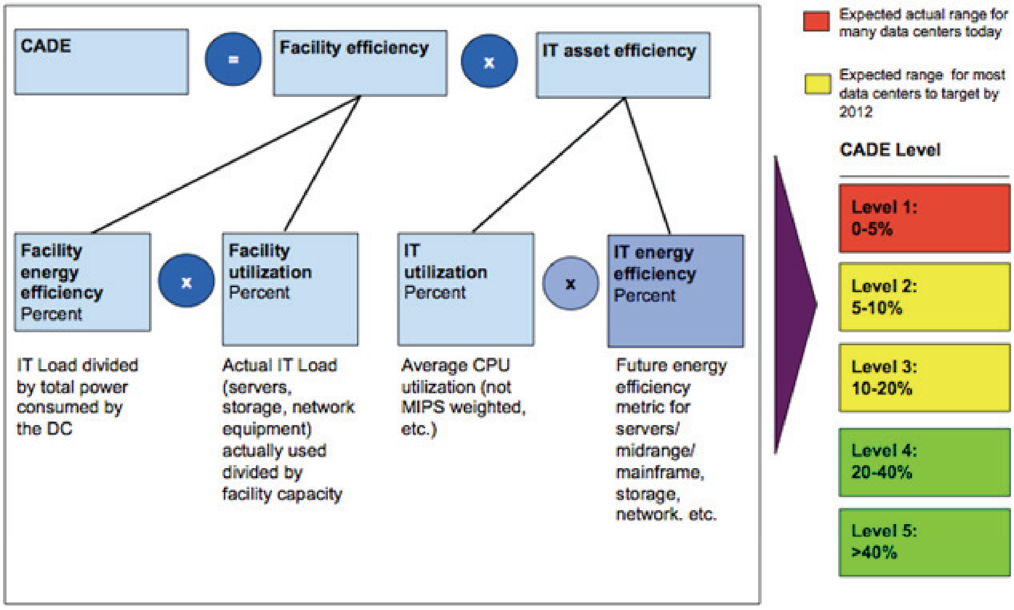


FIGURE 8 Implement metrics for Data Centre Energy Efficiency [20]

The CADE value is relatively straightforward to use and gives a good overall picture of the efficiency of a data centre. The disadvantage of the metrics is that utilization levels are based more on estimates than factual numbers. [18]

2.4.3 *Computer Units per Second*

Computer Units per Second (*CUPS*) represents a proxy for a universal measure of computing output. One Mega CUPS (10^6 CUPS) is equal to the average server performance in 2002. CUPS can serve as the numerator in the equation that determines Compute Efficiency, with the power draw as the denominator: Compute Efficiency = CUPS/Watts Consumed. [21]

Task 1
Calculate the CUPS value of your data centre using Emerson Network Power's calculator at:
http://www.emersonnetworkpower.com/en-ASIA/Latest-Thinking/EDC/Resources/Pages/EnergyLogicEfficiencyCalculator.aspx

3 Process for Data Centre Energy Efficiency

Gaining energy efficiency in a facility can be summarized in a four-step process described in figure 9. Each step is briefly described in a following section.

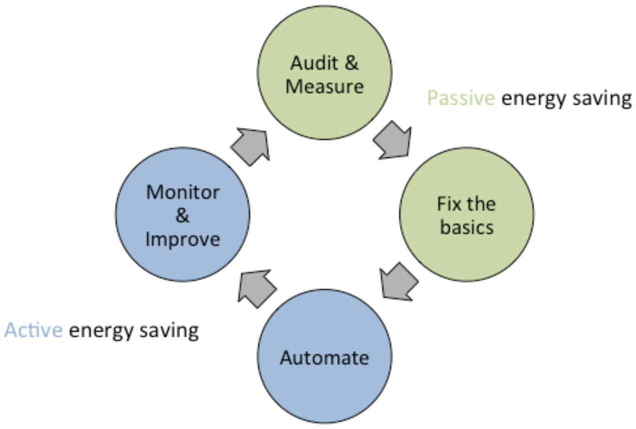


FIGURE 9 Four-step process for improving energy efficiency

3.1 AUDIT & MEASURE

Energy audit

An *energy audit* in a data centre should lead to increased understanding of the current energy consumption, ability to pin point the potential to save energy and a list of prioritized actions. Measurement as stated earlier should be real-time and the minimum long term. The Green Grid recommends using annual measurement results for calculating different xUE levels. Measuring is discussed more thoroughly in Chapter 5.

3.2 FIX THE BASICS

Points of presence

Energy efficiency may be categorized in two: *Passive* and *active* energy efficiency. Passive energy efficiency is related to fixing the basics. Passive energy efficiency does not necessarily need big investments but is more about investing efficient devices, low consumption devices, fixing air flow and so on. In figure 10 are results from Google's five networking rooms (called POP, *points of presence*) units achieved by improvements in passive energy efficiency.

PUE improvements for five POPs

	POP 1	POP 2	POP 3	POP 4	POP 5
Starting point	2.4	2.2	2.2	2.4	2.4
After immediate improvements	2.2	2.0	2.0	2.2	2.0
After cold aisle containment	2.0	1.8	1.8	2.0	1.9
After adding CRAC air return extensions	1.7	1.7	1.7	1.7	1.7
After adding new CRAC controller	1.5	1.5	(Still collecting data)	1.6	(Still collecting data)

FIGURE 10 PUE improvements for five POPs [7]

Without major investments (new CRAC units) Google was able to improve PUE level from 2.4 to 1.7.

3.3 AUTOMATE

Passive energy efficiency is saving energy silently. Active energy efficiency is needed to maximize and sustain the gains from passive energy efficiency. Active energy efficiency means using energy more intelligently to achieve the same results with less. Usually the easiest and quickest way to achieve energy savings is changing policies of an existing environment. This is done by increasing the automation level, add monitoring and maintenance of facilities.

Automation is discussed more thoroughly in Chapters 5 and 6.

3.4 MONITOR & IMPROVE

Savings can be lost without continuous monitoring and maintenance of the environment. Automated systems can drift away from optimal performance as small adjustments are made over time. People can slip back to energy waste behaviour if deviations go undetected and expectations are not reinforced.

Monitoring is discussed more thoroughly in Chapter 6.

Task 2
Check Google's best practices for increased efficiency at
http://www.google.com/about/datacentres/efficiency/external/

3.5 DATA CENTRE MATURITY MODEL

Data Centre Maturity Model (DCMM)

The Green Grid has developed the *Data Centre Maturity Model (DCMM)* to outline capability descriptors by area such that users can benchmark their current performance, determine their levels of maturity, and identify the ongoing steps and innovations necessary to achieve greater energy efficiency and sustainability, both today and into the future. The maturity model touches upon every aspect of the data centre including power, cooling, compute, storage and network. The levels of the model outline current best practices and a 5-year roadmap for the industry. [13]

DCMM is divided in two: *Facility* and *IT*. The facility part consists of four areas: *Power*, *Cooling*, *Management* and *other*. These different areas are measured with levels from 0 to 5 where level 0 is minimal and 5 is highest, visionary level. [14]

Task 3
Take a look at The Green Grid DCMM tool at:
https://www.thegreengrid.org/Global/Content/Tools/DataCentreMaturityModel
and check the areas that are related to data centre facilities (power, cooling, management and other). Try do define the maturity level of your data centre based on DCMM tool scale.

4 Improving Data Centre Cooling and Energy Efficiency

This section summarizes the most common solutions used to improve Data Centre Energy Efficiency with respect to cooling. In the next chapters the improvements with respect IT equipment are discussed.

Mixing hot and cold air flows may cause up to 50% loss in cooling efficiency. [18]. Therefore several techniques to prevent this mixing have developed.

4.1 COLD AISLE AND UNDERFLOOR AIR CONTAINMENT

Cold and hot aisle

cold aisle containment (CAC)

Pervilä et. al. states that both *cold and hot aisle* containment techniques are simple in their key idea: either the hot or the cold aisle is covered at the top and edges of the aisles. This forces the hot and cold air streams to separate. The shaded areas in Figures 11(a) and 11(b) show how *cold aisle containment (CAC)* limits the flow of the cold air stream so that it must pass through the equipment racks. In both cases the aisles must be refurbished so that leakages are minimized. Reasonably airtight doors are required at the edges to allow for operator access, and cable ducts must be isolated to prevent leakages. Different vendors' solutions range from purchasing entirely new racks to installing plastic curtains constraining the flows of air. Obviously, replacing the racks is a very time-consuming and delicate operation, which makes retrofit-capable solutions more desirable. [2]

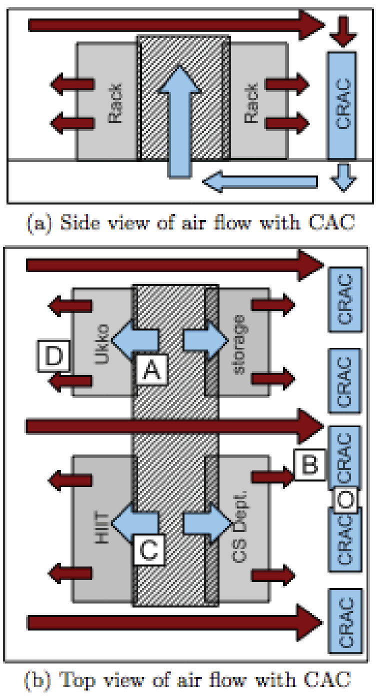


FIGURE 11 Separation of cold and hot aisle streams [2]

Underfloor air containment UAC

The concept of *underfloor air* is not new and initially started in computer rooms at a time when mainframe computers generated considerable heat and had required a considerable amount of complex cabling. Access floor systems allowed plenty of open space to run cabling and a generous pathway to supply large quantities of cooling air under the intense heat of the electronics. The natural convection currents of warm air rising allowed cool air to enter at a low level, cool the equipment and remove the warm air near the ceiling. From the cooling perspective data centres are returning back to the time of mainframe computers. [11]

Drawbacks of UAC

Pervilä et. al. [2] summarizes combined usage of UAC and CAC that in theory, the main *drawback* of installing CAC or UAC is the reduction of supply air volume in the data centre. This means that in case of a power supply failure, there is a smaller reservoir of cold air in the data centre. This flaw must carefully be balanced against the benefits of CAC. The main benefit is that CAC can much more easily be retrofitted into an existing data centre. By comparison, a HAC setup requires considerably more complete air ducts for the exhaust or return flow. As neither HAC nor CAC can entirely avoid overheating scenarios, it is our recommendation that the shutdown temperatures of servers should not be disabled. Fortunately, in most commercial servers this remains impossible.

Another drawback is that UAC may not be applicable in all data centre environments. If the CRAC units are distributed evenly along all of the walls of a data centre, there may not be suitable floor areas for installing UAC. This is an unavoidable problem of some data centre environments. However, for global energy reductions to occur, it is enough that UAC is employed in those cases where it remains applicable. [2]

Finally, UAC is not able to remove or even diminish turbulence caused by underfloor blockages. Despite this drawback, it presents a sizable improvement in air velocity through the perforated tiles in the CAC. A 9% improvement in CRAC blower speed achieved by usage of UAC means that more servers can be installed in the data centre. In addition, Pervilä et. al. have earlier shown [2] that in the same conditions, CAC yielded an improvement of 20%.

CAC and UAC combined can reduce CRAC blower power by 30%

As both CAC and UAC can be installed very cheaply, their combined enhancement of almost 30% CRAC blower power makes the payback time very attractive. CAC is by now a very much standard data centre technique for improving air flow. Based on their studies, Pervilä et. al. [2] hope's that UAC will also catch on.

Figure 12 depicts the combination of CAC and UAC.

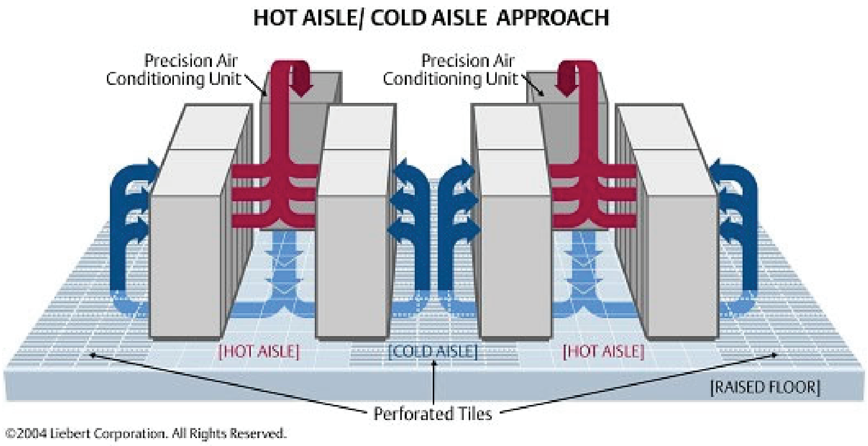


FIGURE 12 CAC and UAC combined for improved CRAC performance [16]

4.2 MODULAR DATA CENTRE

The concept of a Modular Data Centre solution has eluded definition, if not comprehension. Through the short history of modular solutions and vendor marketing, a definition and categorization of solutions has emerged. The word 'modular' means a self-contained unit or item that can be combined or interchanged with others like it to create different shapes or designs. [12]

Definitions

More specific, data centre oriented definition for a container and modular are:
– *Container* is a data centre product incorporating customized infrastructure to support power or cooling infrastructure, or racks of IT equipment. Containers are built using an ISO intermodal shipping container.
– *Modular* is an approach to data centre design that implies either a prefabricated data centre module or a deployment method for delivering data centre infrastructure in a modular, quick and flexible method. [12]

The concept of Modular Data Centre has been around over a decade. Around 2002 Google began experimenting with a container full of IT. APC (currently Schneider Electric) took an early approach to modularizing the data centre in 2004. A product called InfraStruXure Express was a mobile data centre truck with integrated power, cooling and racks. The primary use for it was disaster recovery and temporary or transitional IT projects. In January 2007 Sun Microsystems introduced the Blackbox data centre container. It is considered a kick-start for the notion of a Modular Data Centre. Google and Microsoft have continued to innovate their own solutions into a Modular Data Centre design. [12]

Modular Data Centre solutions are ideal for both green- and brownfield locations. It is a technology worth to consider when making new data centre investments.

CSC – IT Centre for Science Ltd. is aiming to build one of the most eco-efficient data centres in the world. The location is Kajaani, in Northern Finland. Datacentre CSC Kajaani is a proven solution based on technology, modern, reliable infrastructure and ecological efficiency for data needs in research and development in public and private sector. CSC's data centre is a Modular Data Centre and started in production use in 2012. CSC's data centre aims for a near zero carbon footprint during its operational lifetime. Servers are cooled down with outside air and the primary means of electric power is hydropower-generated energy. The annual PUE of the data centre is estimated to be 1.15. [8]

CSC's chosen technology for Modular Data Centres is SGI Ice Cube Air R80 which promises up to 1.06 PUE for its products [10]. SGI module installed at Kajaani may be seen in figure 13.

FIGURE 13 CSC's Modular Data Centre at Kajaani, Finland

Task 4
Take a look at a video of building a Modular Data Centre. Fast forward the video to *47:30* for two minutes part. Video may be found at
http://qsb.webcast.fi/c/csc/csc_2012_1016_csc/#/stream

4.3 APPROVAL FOR ENERGY EFFICIENCY PROJECTS

As an IT administrator it is crucial also understand the logics behind business decisions. Advanced technology itself rarely is a reason for extra investment but business decisions should be based on earning more money out of the investments. Below are some factors that drive decision making and are worth to remember when reasoning energy efficiency projects in your organisation:

1 Present other projects within same framework
2 Include applicable rebates and incentives
3 Emphasize reduced maintenance costs and savings over time (Return of investment, *ROI*)
4 Simple payback
– How long it will take for an investment to begin making money (Internal rate of return, IRR)

Decreased energy consumption equals smaller electricity bill but energy efficiency also has marketing value for the company. Green values are increasingly recognized by consumers and may affect their purchase behavior and decision-making. Business decisions are discussed more thoroughly in Chapter 5 and 6.

5 Tier classification

Classify the reliability

In section 1 we discussed the configuration of a data centre. The specific components which have to be present depends on how reliable the data centre has to be. To *classify the reliability*, the Uptime institute defined a Tier system[1]. The following list summarizes the high level characteristics of each tier.

Tier 1

Tier 1 is composed of a *single* path for power and cooling distribution, *without* redundant components, providing 99.671% availability.

Tier 2

Tier 2 is composed of a *single* path for power and cooling, *with* redundant components, providing 99.741% availability.

Tier 3

Tier 3 is composed of *multiple* active power and cooling distribution paths, but only one path is active, has redundant components, and is concurrently maintainable, providing 99,982% availability.

Tier 4

Tier 4 is composed of multiple active power and cooling distribution paths, has redundant components and is *fault tolerant*, providing 99.9995% availability.

It is obvious that the more redundancy and fault tolerant the more expensive the data centre management is.

SUMMARY

In this chapter the structure of energy consumption of a data centre was presented. Biggest influence to the energy efficiency (or lack of it) of a data centre in most cases is cooling. Basic physics of heat, humidity and pressure – which affects the cooling – were introduced. Results of big data centre operators' studies and best practices based on those experiences to reduce energy consumption were listed and briefly presented. Below is a summary by Barroso et. al. [1] of the practices to reduce energy consumption:

1 Careful air flow handling: segregate hot air exhausted by servers from cold air, and keep the path to the cooling coil short so that little energy is spent moving cold or hot air long distances.

2 Elevated temperatures: keep the cold aisle at 25-30°C rather than 18–20°C. Higher temperatures make it much easier to cool data centres efficiently. Virtually no server or network equipment actually needs intake temperatures of 20°C, and there is no evidence that higher temperatures cause more component failures.

3 Free cooling: in most moderate climates, free cooling can eliminate the majority of chiller runtime or eliminate chillers altogether.

[1] Turner, W.P., Seader, J.H. and Brill, K.G. 2005, Industry Standard Tier Classifications Define Site Infrastructure Performance, Uptime Institute, white paper

4 Better power system architecture: UPS and power distribution losses can often be greatly reduced by selecting higher-efficiency gear, as discussed in the previous chapter.

Different metrics – some well-known and less so – were introduced and explained briefly. Finally tools and processes were suggested to improve the energy efficiency of a data centre.

Literature

[1] Barroso et. al., The Datacentre as a Computer, http://www.morganclaypool.com/doi/abs/10.2200/S00516ED2V01Y201306CAC024

[2] M. Pervilä et. al., Data Centre Energy Retrofits, Helsinki University 2013

[3] Schneider-Electric, Schneider Energy University, http://schneideruniversities.com/energy-university/

[4] C. Malone and C. Belady. Metrics to characterize datacentre & IT equipment energy use. ®In Proceedings of the Digital Power Forum, Richardson, TX, September 2006.

[5] The Green Grid, WUE white paper, http://www.thegreengrid.org/~/media/WhitePapers/WUE

[6] The Green Grid, CUE white paper, http://www.thegreengrid.org/en/Global/Content/white-papers/Carbon_Usage_Effectiveness_White_Paper

[7] Google, Google's Green Data Centres: Network POP Case Study, http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/corporate/datacentre/dc-best-practices-google.pdf

[8] CSC, Datacentre CSC Kajaani homepage, http://www.csc.fi/english/datacentre

[9] CSC, Datacentre CSC Kajaani opening ceremonies, http://qsb.webcast.fi/c/csc/csc_2012_1016_csc/#/stream

[10] SGI, Ice Cube Air white paper, http://www.sgi.com/products/data_centre/ice_cube_air/

[11] Carrier, Design guide to underfloor air distribution, http://www.docs.hvacpartners.com/idc/groups/public/documents/techlit/ds_ufad.pdf

[12] Rath Consulting, Data centre knowledge guide to Modular Data Centres, http://whitepapers.datacentreknowledge.com/?option=com_categoryreport&task=viewabstract&pathway=no&autodn=1&title=15932&crv=0&src=3956&ctg=655&cmp=4101&yld=1

[13] The Green Grid, DCMM tools homepage, https://www.thegreengrid.org/Global/Content/Tools/DataCentreMaturityModel

[14] The Green Grid, DCMM poster, https://www.thegreengrid.org/Global/Content/Tools/~/media/Images/DCMM/DCMM_StandardE.pdf

[15] Kothandaraman, C.P., Fundamentals of Heat and Mass Transfer, New Age International 2006

[16] Core Airconditioning Ltd., Company homepage, http://www.coreac.com/airconditioning/products/computerroom.htm

[17] The Green Grid, PUE white paper, http://www.thegreengrid.org/sitecore/content/Global/Content/white-papers/The-Green-Grid-Data-Centre-Power-Efficiency-Metrics-PUE-and-DCiE.aspx

[18] Motiva, Energiatehokas konesali, http://www.motiva.fi/files/4828/Energiatehokas_konesali.pdf

[19] Greijer, A., Energieffektivitet i datahallar, http://www.eko-fysisk datormiljo.se/ovrigt/NPUE.pdf

[20] McKinsey & Company, Revolutionizing Data Centre Energy Efficiency, http://www.ecobaun.com/images/Revolutionizing_Data_Centre_Efficiency.pdf

[21] Emerson Network Power, Energy Logic: Calculating and Prioritizing Your Data Centre IT Efficiency Actions, http://www.emersonnetworkpower.com/documents/en-us/latest-thinking/edc/documents/white%20paper/energylogicmetricpaper.pdf

[22] Viawest, PUE: The Measure of Accountability in Data Centre Power Use industry white paper, http://www.viawest.com/sites/default/files/asset/document/PUE_-_The_Measure_of_Accountability_in_Data_Centre_Power_Use.pdf

[23] Wikipedia, Power Distribution Unit, http://en.wikipedia.org/wiki/Power_distribution_unit

Barroso, L.A., Clidaras, J. and Hölzle, U. The Datacentre as a Computer. An Introduction to the Design of Warehouse-Scale Machines, Second Edition, Morgan & Claypool Publishers, 2013.

Pelley, S., Meisner, D., Wenisch, T.F., & VanGilder, J.W. (2009). Understanding and abstracting total data centre power. Paper presented at the Workshop on Energy-Efficient Design

Pervilä M. Data Centre Energy Retrofits, University of Helsinki, 2013.

Salim, M. (2009). Energy In Data Centres: BENCHMARKING and Lessons Learned. (Cover story). Engineered Systems, Volume 26 (Issue 4), 24-32.

MODEL ANSWERS

Answers to Reflection Questions

1 According to ASHRAE data, cited by Green Grid, the recommended operating temperature for IT equipment is in the range 18 – 27 degrees Celsius[2]. During its normal operation, IT equipment produces large amounts of energy. Heat cannot be destroyed so it needs to be channeled out of a data centre in one way or another. Also the more controlled the process is the better. For example conduction and convection occur when heat is conducted away from processor and other computer parts through heat sinks into the surrounding, flowing air. Regulating air flow and keeping it at optimal temperature and separated from warm air increases the efficiency of cooling.

2 There should be a balance between enough moisture to prevent static electricity (can damage the IT equipment) and not too much moisture (than condense forms on the IT equipment; with short circuit and fire). *How part* is data centre related. Usually there are own equipment for that or it is done with CRAC units. Important it is because of static electricity though risk for that is in minor role. Dew point then again is one way of controlling co-operation of different CRAC units of a data centre.

3 They all are telling the same thing but from different approach. Same variables exist in different equations and when used together along the recommendations of The Green Grid will give a realistic overview of a Data Centre Energy Efficiency.

4 From some perspective 'rules' for calculating a PUE value are too loose which causes intentionally or unintentionally wrong PUE values. This may take away the credibility of a PUE value as a metric. Then again before PUE there were no unified, standard way to measure and compare data centres energy efficiency at all. And introduction of a PUE metric was one of the first steps towards comparable energy efficiency measurement.

5 No single, perfect metric exists. New metrics come from a need that earlier metrics do not fulfill. On the other hand the authority coming up with metrics wishes to define the way energy efficiency is measured since it brings business for the authority at issue. More metrics bring scale for the subject but may reduce the comparability of different results if all parties do not follow certain metrics. Certain simplicity keeps comparing easy and understandable for common sense but at the same time could subordinate metrics for malpractices.
In order to make improvements comparable this course uses the EU Code of Conduct on Data Centres as the guideline.

[2] https://www.thegreengrid.org/~/media/WhitePapers/WP50-Data%20Center%20Efficiency%20and%20IT%20Equipment%20Reliability%20at%20Wider%20Operating%20Temperature%20and%20Humidity%20Ranges.pdf?lang=en