TNO innovation for life

Energy flexibility interfaces

IPCEI-CIS MISD - Deliverable 2.7
Version 1.0

TNO Public ⟩ TNO 2025 R10799
30 June 2025

ICT, Strategy & Policy
www.tno.nl
+31 88 866 70 00
info@tno.nl

TNO 2025 R10799 – 30 June 2025

Energy flexibility interfaces

IPCEI-CIS MISD - Deliverable 2.7
Version 1.0

Author(s)	Bram van der Waaij (TNO), Mente Konsman (TNO)
Reviewer(s)	Maria Vlasiou (UT-EEMCS), Marco Gerards (UT-EEMCS), Magiel Bruntink (TNO)

Title	TNO Public
Number of pages	36 (excl. front and back cover)
Programme name	IPCEI-CIS

This activity is conducted in the scope of the Modular Integrated Sustainable Data Center (MISD) project that received funding from the Netherlands Enterprise Agency (RVO) under the 8ra initiative (IPCEI-CIS).

Contents

1 Introduction

The energy transition fundamentally changes the way consumers (small and large) interact with the electricity system. In the past, consumers could simply count on electricity producers to match their production with the energy needs of the consumers. The electricity grid also had sufficient capacity to deliver the electricity around the clock. The needs of the consumers were met by the electricity system and they could afford to play a rather passive role.

Due to the changes brought about by the energy transition, consumers are increasingly becoming an active part of the electricity system. They are being asked to be more flexible in their consumption patterns. Since data centres are major consumers of electricity, the question arises what impact these developments will have on them. This can be divided into two sub-questions:

1. Where are the opportunities for data centres with regard to energy flexibility?
2. Where should this flexibility come from?

This deliverable investigates the potential of energy flexibility in and for data centres and designs the necessary interface(s) between the relevant stakeholders.

Why is flexibility becoming a relevant topic for data centres? From an energy perspective: what are the issues around sustainable energy and how do they have an impact on data centre operation? Which NL/EU legislations are already in place and are expected to be installed? How does legislation influence the data centre operation and perhaps even their customers now and in the upcoming years? What should (new) interface look like between the energy operators and the data centres, and between the data centres and its customers?

Within the context of this deliverable making data centres more sustainable means reducing CO2 emissions. Other relevant sustainability topics, such as embodied carbon, are out of scope because their possible contribution to flexibility is low.

1.1 Trias Energetica

Reducing CO2 emissions of data centres boils down to reducing their consumption of *grey* energy . A practical strategy to reason about this is the trias energetica[1].

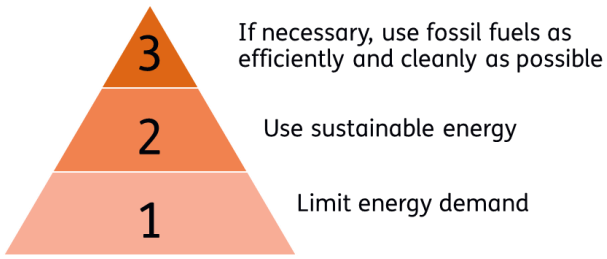

Figure 1: Trias energetica

The three steps of the trias energetica are the main rules for designing sustainable electrical systems:

1. **Limit the use of energy.**
   Determine if a particular task is really necessary to perform. Not doing something saves typically energy and is therefore a very sustainable solution. Also avoid wasting energy by for instance stopping machines that are not used.

2. **Use sustainable energy**.
   In those circumstance that you cannot limit the use of energy, use as much sustainable energy as possible, like solar and wind[2].

3. **If necessary, use fossil fuels as efficient and clean as possible.**
   This is outside of scope of this deliverable

Flexibility comes into play in step 2: By becoming flexible more sustainable energy can be used. Becoming flexible for instance means being able to deal with periods green electricity is not available.

[1] Trias energetica - Wikipedia
[2] It is debatable if nuclear is a clean energy source, but it does not emit CO2.

1.2 Flexibility versus Efficiency

To understand the meaning of flexibility it is useful to compare it with efficiency. The table below compares efficiency and flexibility from a number of different viewpoints.

Table 1: Efficiency versus flexibility

	Efficiency	Flexibility
Overhead	Min possible overhead	Overhead by design
Advantage	Direct advantage for the data centre	Advantage for external party (energy stakeholders). Incentives are needed to convince the data centre
Functionality	Keep full functionality	Two options to create flex space: • Data centre accepts additional resource requirements: *Full customer functionality with additional overhead* • Customers accept lower SLA levels for (financial) benefits: *Reduced customer functionality with no overhead*
Design time	• Optimize algorithms • Deploy minimal resources to provide full functionality	Determine amount of additional resources and/or acceptable SLA levels to create room for flexibility (financial balance between primary process and providing flex).
Run time	• Dynamic optimizing resources for lowest costs. • Keep full functionality • No impact on customers	Continuous responds to flex requests.

It is important to realize that both efficiency & flexibility are utilizing the same underlying technologies. The technology is not different, the reason and moment it is used is different.

From the trias energetica perspective efficiency fits perfectly step 1: limit energy usage. Being more efficient means using less energy[3]. Flexibility on the other hand can be used in both steps 1 and 2. Flexibility enable the of use green energy more efficiently.

Table 2: Efficiency and flexibility in the Trias Energetica

	Efficiency	Flexibility
1) Limit energy usage	X	X
2) Use sustainable energy		X

[3] Depending on the used definition of effiency, but in this deliverable they are the same

1.3 Readers guide

The next chapter, chapter 2, explains if and when flexibility is relevant for data centres. Chapter 3 goes into detail what relevant legislations in the NL and the EU are already in place and currently in development. Both chapters explain the upcoming inevitability of flexibility in data centres.

Then chapters 4 and 5 provide an introduction to the data centre - and the energy domain for readers that are not familiar with one or both worlds.

Chapter 6 investigates the flexibility capabilities of data centres. Chapter 7 uses that knowledge to discusses the different interfaces needed to incorporate electricity flexibility in the data centre and how to involve its customers.

Finally, the conclusions section (chapter 8) will formulate answers to the questions posed in this introduction.

2 Why flex for data centres?

This chapter investigates if and when energy flexibility (both electricity and heat) is relevant for data centres. Reasons for offering flexibility are discussed based on a number of sustainability trends that impact data centres.

2.1 Trends

Looking at the current state of energy availability and the desire to be sustainable a couple of trends can be identified:

- Limited electricity grid capacity
  Easily obtaining a grid connection is no longer a given. Especially in the Netherlands more and more limitations are added on the electricity grid. There are long (sometimes years) waiting lists for obtaining new electricity connections or upgrade existing ones at a lot of locations in the Netherlands. These leads to discussions about which types of companies/organizations should be prioritized to be connected to the grid first. Currently data centres are not at the top of that list.

- Each year CO2 emissions must be lowered
  EU regulations state CO2 emissions must be reduced further each year. This also applies to electricity generation leading to an ever smaller share of grey electricity. By 2050 emission should be zero, which also implies 100% green electricity production.

- More sustainable heating
  Heating is important for housing in the winter period and for some industry. From the sustainability drive there is a strong incentive to become less dependent on fossil fuels. A search for alternative heating sources is going on.

  Data centres are an interesting candidate because they produce a lot of heat, although currently there are still several mismatches:
  - the temperature of the produced heat by data centres is typically low (around 27 ^{0}C)[4], while the desired temperature for heating of houses and industry is typically higher (around 70 ^{0}C)[5]. Heat pumps are needed to bridge this gap, making the solution less sustainable.
  - the amount of heat produced by a single data centre does not match the local demand for heat. Large data centres produce much more heat than needed for the local heating of houses, resulting in (far more) additional cooling capacity making the solution less sustainable.
  - between periods heat is needed. In non-industrial settings, such as heating houses, heat is typically needed more in the winter than in the summer.

[4] Hoe blijft een datacenter gekoeld tijdens een hittegolf? - Dutch Data Center Association
[5] Warmtenetten: hoe ze werken, waarom ze nodig zijn en hoe ze worden gerealiseerd | Nationaal Warmte Congres

The combination of these trends together with the expectation that the demand for data centre capacity will keep increasing in the coming years[6] (partly due to the exponential growth of AI), it is obvious that data centres have a challenge to deliver the desired capacity in a sustainable manner.

2.2 Why do you need flexibility?

Solutions that make data centres and (AI) applications more efficient are essential, but not all problems can be solved with only efficiency measures. Some problems do not occur continuously, but are time and/or location dependent:

- Congestion in the electricity grid
  When in the electricity grid congestion occurs it is typically in specific areas and only occasionally. For instance at really sunny moments PV panels produce a lot of electricity, in not all areas the local grid might have enough capacity to transport all that electricity. Temporal additional local consumption can help reduce this problem.

- Availability of green electricity
  The availability of green electricity is not guaranteed, it is not a continuous source. Windmills only work when there is wind, PV panels only produce electricity when there is sun. While they are both somewhat predictable, they are certainly not always available. Being able to continuously adapt electricity use to the availability of green electricity will help to reduce this problem.

- Locality of green electricity
  In addition to the fact that green energy is not always available, the location where that green energy is available also varies. After all, where the wind blows or the sun shines varies.

- Security of supply
  In the coming years the certainty of electricity will go down in the Netherlands. Tennet states in its annual report[7]: "Until 2030, the security of supply in the basic forecast remains within the established standards. But after that, the risk of shortages increases rapidly".

By responding flexibly in time, duration, volume and/or location to these four problems, data centres can help making the electricity grid more stable and at the same time becoming themselves more sustainable. Flexibility technologies can help data centres to become more sustainable by making optimal use of their resources regarding energy usage.

Next to electricity also heat is sometimes seen as a flexibility resource, but not in this deliverable. In most cases heat demand is continuous or changes rarely over time (in months or seasons). Industrial process that need heat, need heat often all year round. Home heating is seasonal. Mitigating seasonal changes would require enormous buffer capacity, which in the Netherlands is only realistic using geothermal energy. Another possible solution is to create a specific combined set up. For instance in combing heat demand in the winter and solar energy in the summer would be an interesting combination to become overall more sustainable. For data centres that is sometimes an option,

[6] McKinsey states "that global demand for data center capacity could rise at an annual rate of between 19 and 22 percent from 2023 to 2030".
[7] https://www.tennet.eu/nl/over-tennet/publicaties/rapport-monitoring-leveringszekerheid

depending on whether green energy is produced nearby and what (seasonal) heat demanded is. Both heat demand solutions do not require flexibility solutions.

2.3 Outlook

What are the expectations for the coming years?

- Congestion in the electricity grid
  The expectation is that the congestion problems in the Dutch electricity grid will remain for at least 10 – 15 years.

- Expansion of data centres
  Due to the ongoing problems in the electricity grid, Dutch data centres cannot easily expand. But especially in the light of the exponential growth of AI expansion is demanded.

- CO2 emissions
  The emphasis on CO2 emission reduction will not decrease but rather increase in relation to the EU target of CO2 neutrality in 2030

2.4 Summary

Sustainability is already an important topic for data centres. Due to the expected developments in the coming years on Dutch grid stability and EU CO2 emission reduction, in the long run flexibility is inevitable. As a result, future data centres in must become more and more flexible in their energy usage.

The remaining chapters in this report look at how data centres can continue to meet this increasing demand for flexibility.

3 Policy drivers

This chapter focuses on policies that will incentivize data centres (and other consumers for that matter) to become more energy flexible over time. Most of the discussed policies are included in the EU Fit for 55 package, which bundles a lot of the EU policies and legislation that are aimed at reducing CO2 emissions.

3.1 EU Fit for 55 package

The Fit for 55 package is a set of laws that was adopted by the EU council in 2023 and stipulates objectives for reducing CO2 emissions in Europe. The package name stems from the target to reduce CO2 emissions by 55% by 2030 , compared to the emissions level in 1990. This is a significant step up in reductions compared to the previous goals for 2030. Ultimately, in 2050, the CO2 emissions should go down to zero.

3.2 EU Emissions Trading System

One of the main mechanisms to realize the Fit for 55 reduction targets is the EU Emissions Trading System (EU ETS). The EU ETS covers the following sectors: electricity and heat generation, energy-intensive industries and commercial aviation. Although data centres are not directly affected by the ETS, there is an indirect impact via electricity generation.

The ETS revolves around allowances to emit one tonne of CO2 or the CO2 equivalent of that in terms of other greenhouse gasses. Each year, companies in sectors that are covered by the ETS need to buy a number of allowances that matches their greenhouse gas emissions. However, there is a cap on the number of allowances; in 2024 that number equated to 1,386,051,745.

Each year this cap will be reduced, which gradually drives up the price of allowances. This creates an incentive for companies to lower their emissions. The yearly reduction percentage is 4.3% for the 2024-2027 period and 4.4% between 2028 and 2030.

As already mentioned, data centres will only be affected indirectly by the EU ETS through the effects it has on electricity generation. The ETS will lead to a reduction of grey electricity in the electricity mix.

A data centre can already opt to use green electricity only. This can be proven using guarantees of origin (GO's) bought from qualified renewable electricity producers. The total amount of electricity consumed by a data centre should match the amount of electricity covered by the GO's that were purchased. The electricity consumption and renewable production typically only have to add up over a longer period of time, such as a year.

With this current system of GO's one can "prove" that electricity consumption was 100% green during the whole year. However, this is only a paper reality as this does not represent the real green electricity consumption.

Paradoxically, the more grey electricity production there is in the system and the lower the demand for GO's, the easier it is to compensate for hours with insufficient renewable production and still be 100% green on paper. When there is insufficient green electricity available, one can simply use grey electricity instead. This grey electricity consumption can be made green afterwards by buying a corresponding number of GO's. This will become much harder when the demand for GO's will be higher and the share of grey electricity lower. Ultimately, this means that for periods of low availability of renewable production, electricity consumption needs to be reduced or consumption will have to come from previously stored green electricity. However, it may take some time before the reduction of the emissions caps will have a significant impact on the actual electricity consumption patterns of data centres.

There is another trend that could create that impact a lot sooner than just the lowering of the EU ETS caps alone. As already discussed, the current guarantees of origin certificates do not truly reflect how much green electricity is being consumed. This is also being acknowledged by regulators and legislators who are looking into hourly matching to make the system fairer.

3.3 Hourly Matching

Hourly matching simply means that the amount of green electricity that is consumed within a certain hour has to be matched by an equal amount of renewable production in that same hour.

An important step towards hourly matching has already been made by updating the guarantees of origin in the Renewable Energy Directive (REDIII) in 2023, which is also part of the EU Fit for 55 package. Originally, GO's were labelled with a timestamp that only indicates the month and year of production, but now they can be much more fine grained; up to 15 minute intervals (corresponding to the imbalance settlement period). The availability of such high resolution timestamps is an important prerequisite to implement hourly matching.

3.3.1 24/7 Carbon Free Energy

Although there are already some elements in place that could facilitate hourly matching in the EU, a legislative framework is still to be developed.

In 2021, the United Nations launched their "24/7 Carbon-free Energy Compact to Accelerate the Decarbonization of Electricity Grids". The objective of this call to action is described as follows:

> *"Full-scale electricity decarbonization fundamentally means that every kilowatt-hour of electricity demand is served by carbon-free electricity sources, every hour of every day, everywhere."*

To help realize this objective the 24/7 CFE is based on the following five principles: time-matched procurement, local procurement, technology-inclusive, enable new generation and maximizing system impact.

Signatories of the 24/7 CFE include big tech companies Google and Microsoft.

There is also a growing number of utility companies, e.g. Engie and Vattenfall that offer 24/7 CFE services, including reporting platforms and hourly matched energy management services.

3.4 EU reporting on sustainability of data centres

The European Energy Efficiency Directive (EED), which in its latest revision is also part of the Fit for 55 package, sets out goals for energy efficiency. Its main target is a reduction of energy consumption in the EU of 11.7% by 2030, compared to 2020. It sets out regulations and reporting requirements for different sectors such as the public sector and public buildings.

There is also a requirement in the EED to report on the energy performance and sustainability of data centres. This requirement has been worked out in a concrete reporting scheme (Delegated Regulation (EU/2024/1364)[8]). One of the aspects that data centres have to report on is the total renewable energy consumption as shown in Figure 2.

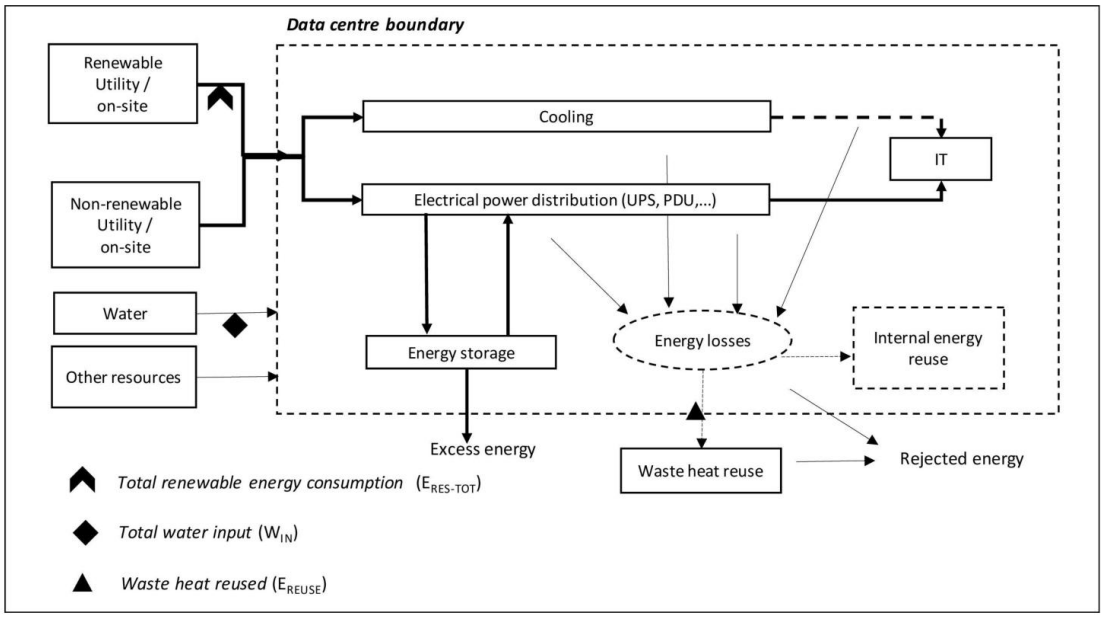

Figure 2: Overview of the measuring and monitoring points in a data centre (taken from Delegated Regulation (EU/2024/1364))

These reports can be used by data centres clients to compare data centres based on their sustainability performances. The reporting period covers a calendar year and also partly relies on guarantees of origin for determining the total renewable energy consumption. Although this reporting scheme is a good step forward in creating more transparency for data centre clients, it suffers from the same issues as the current GO's in that renewable energy consumption is not considered in real-time.

[8] https://eur-lex.europa.eu/eli/reg_del/2024/1364/oj

3.5 Summary

As is apparent from this overview of policy drivers, energy flexibility will become increasingly important to meet the CO2 reduction targets. The gradual reduction of the share of fossil energy and the expected future introduction of hourly matching of renewable energy production and consumption will make it harder and harder to just be green on paper. Over the years, consumption patterns of data centres will have to adapt in real-time to the availability of renewable production and/or bridge these gaps by energy storage solutions.

4 Data centre domain

This chapter gives an overview of what a data centre is and what components it is made of., including the different types of data centres and how are they interconnected. This chapter is intended for those readers that are not familiar with the data centres.

4.1 The basics

The main goal of a data centre is to provide a shared environment to perform compute tasks. Data centres excel in high availability, which is ranked in 4 tiers expressing minimal uptime and maximum downtime levels. Tier 1 is 99.671% uptime per year, which results in a downtime per year of less than 28.8 hours. Tier 4 is 99.995% uptime per year, which results in a downtime per year of less than 26.3 minutes.

For a data centre to provide a shared environment both compute and non-compute equipment is needed. Figure 3 gives an overview of the basics of a data centre, its customers and the connections it has with external infrastructures. The next sections go into more detail on the compute and non-compute components of a data centre.

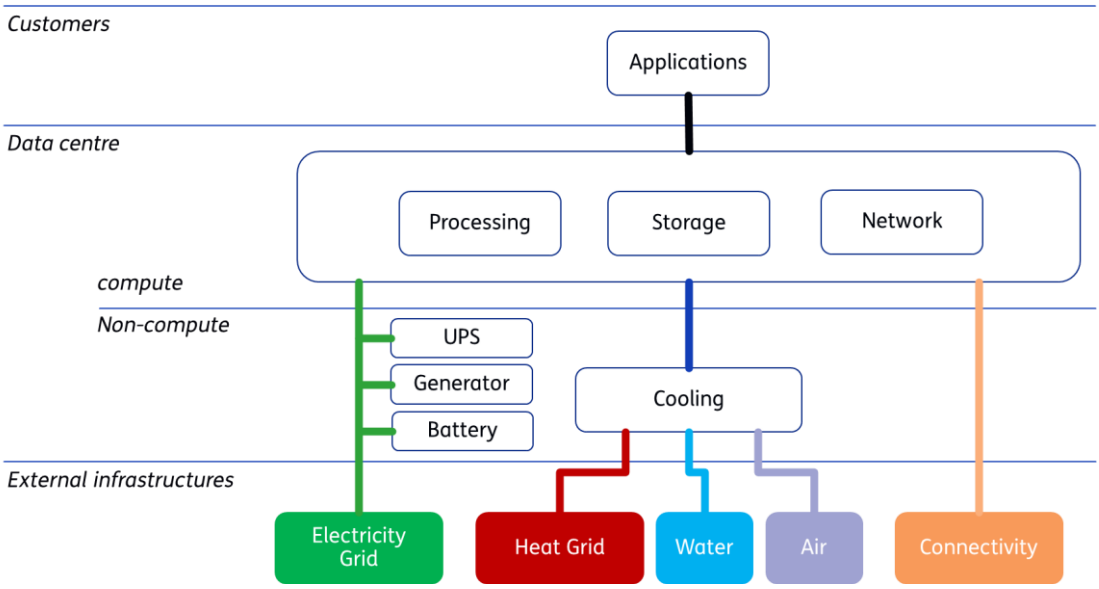

Figure 3: Data centre basics

4.1.1 Compute

In order to provide a shared environment to perform compute tasks, a data centre needs equipment / hardware for processing: computers - often called servers. Beside that a data centre also needs storage and network equipment to interconnect all computers and storage. The network is also important to connect the data centre with the outside world, often the Internet, but also special direct connections to other data centres are common. Summarizing compute consists out of: processing, storage and network equipment. That are all consuming electricity.

4.1.2 Non-compute : cooling

Beside compute equipment also non-compute equipment is needed. Cooling is a very important topic. Computers are producing a lot of heat that must be removed to avoid hardware failures. There are several methods for cooling, the three major types are:

- Air cooling is common, blowing cold air through the hardware and cooling the hot air using special cooling units outside the data centre.
- Water cooling pumping water along the hot chips in a closed circuit with a heat exchanger to cool this water. The heat exchanger can be cooled using cooling units outside the data centre, or by exchanging water with a nearby water source, e.g. a river.
- Liquid immersing cooling putting the entire hardware boards (without the casing) in a liquid bath, often a type of oil. A heat exchanger is in place to cool the oil and evaporate the heat outside the data centre.

An upcoming technology is heat re-use by customers in need of heat. Heat customers can be industrial processes, households and buildings, or any other heat users. Also a heat grid can be used, a more generic approach bringing heat producers and heat consumers together.

4.1.3 Non-compute: maintaining uptime

Beside cooling also other non-compute equipment is relevant for data centres. Another important piece of non-compute equipment is tasked with maintaining uptime by bridging power outages for as long as possible. For this a two-step approach is used:

- Uninterruptible Power Supplies (UPS). These are batteries specifically installed to overcome short periods of power shortage. A battery can take over instantly when power is lost and keep all the hardware running. The capacity is typically enough to bridge 5-10 minutes gaps under full load of the data centre. Which is enough to go to step two:
- Generators. To bridge a longer period of no electricity the data centre must produce its own electricity. This is typically done using diesel generators. As long as there is diesel there is power.

Power is very important for uptime, all the hardware depends on it. But data centres take more measures like spares and duplication. Having spare hardware available that can be switched to quickly when there is a hardware failure. Having more than one network connection to the outside world. Having a double connection with the electricity grid. The double network and electricity is to deal with problems in the outside world such as a cable (network / power) accidentally being cut.

4.1.4 Non-compute: sustainability

A third group of non-compute equipment is emerging and focusses on sustainability. At the moment batteries are the key element and are used to store electricity when needed. Batteries for instance can be charged at moments when there is a lot of green electricity available. By discharging the batteries (and using it in the data centre) during periods of low green electricity in the grid, the overall green electricity usage of the data centre can be enhanced. In order to bridge hours or even days with only batteries, very large batteries are needed. They are still very expensive and therefore at the moment economically not viable for this purpose.

Note that data centres already have large amount of batteries for their UPS functionality. But this is a very different functionality that uses batteries as a risk mitigation against power outage. UPS functionality is very important for data centres, it helps in maintaining a high availability, and can therefore not used for sustainability purposes. Additional batteries must be installed for that.

4.2 Data centre types

Data centres come in various types and sizes. The Dutch Data Centre Association[9] divides them into five categories:

Table 3: Data centre categories

	Size	Energy
Enterprise	~10M^2	0.01 – 10MW
Regional	~200M^2	0.5-10MW
National	~2000M^2	1-10MW
International	~5000M^2	> 5MW
Hyperscaler	~5Ha	> 50MW

Enterprise data centres are used to house the IT equipment of the company. These are typically used for internal purposes and/or for services to their customers.

A hyperscale data centre is very large scale and hyperscale companies have many of them around the globe. Their purpose is to provide services around the world. Currently 7 companies are called hyperscalers: Alibaba, Amazon, Apple, Google, Meta, Microsoft, Tencent. They provide one stop shop, very well integrated advanced hosting and software services as well as their own end user services.

In between are the regional, national and international data centres of other companies. This is a wide variety of data centres of all sizes and build to host compute and (advanced) software services. These data centres services multiple customers at the same time and offer typically no end user services. That is up to their customers.

[9] Wat is een datacenter? - Dutch Data Center Association

4.3 Cloud versus edge computing

Two another important concepts that are relevant to understand are cloud and edge computing and their relation to data centres.

Cloud computing is the delivery of digital services via the internet. Instead of owning and maintaining data centres and servers, companies can access these resources on-demand from a cloud service provider and create their own services and applications based on them. This model offers benefits such as cost savings, scalability, and flexibility. So data centres are used to offer cloud services, which in their turn offer digital services.

Edge computing is a distributed computing paradigm that brings computation and data storage closer to the location where it is actually needed, enabling faster data processing and real-time insights. The main reasons for this are: low latency, high bandwidth and data locality. Edge computing is particularly useful for applications requiring low latency data processing, like autonomous vehicles and gaming. But also the processing of privacy sensitive data, for instance patient data, is often processed at an edge located within a hospital. So data centres can also be used to offer edge computing. If that is the case depends on the location of the data centre relative to its customers, or most likely to the customers of the customer. The actual users of the digital services running on the edge. For instance a gaming company guarantees low latency gaming by making sure it has data centres close to all its customers and on each of these locations it runs a gaming platform to host the customers close to that physical location.

There are a lot of different definitions of the "edge" concept. They differ in the location of the edge data centre. Some people see a small device, such as a mobile phone, as the edge. Others compute on premise such as within the house. But also small data centres across the country are sometimes called edges. The EU has written a report[10] "European industrial technology roadmap for the next generation cloud-edge offering" presenting a nice overview to get an impression, see the figure below.

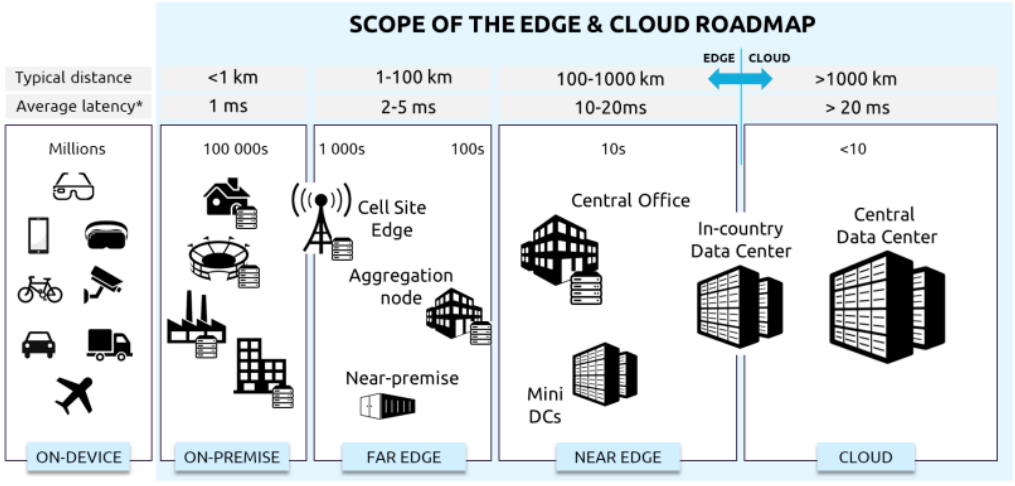

Figure 4: Edge versus cloud computing

More information can be found in the Internet, among others at the *GeeksforGeeks* website[11].

[10] European_CloudEdge_Technology_Investment_Roadmap_for_publication
[11] What Is Cloud Computing ? Types, Architecture, Examples and Benefits | GeeksforGeeks

4.4 Type of compute services

A data centre sells compute services to its customers. There are many different types of compute services, but they can be divided into three main groups:

- Co-location
  These are for customers that have their own hardware, their own servers and storage, that they want to host in a data centre which provides guaranteed power, cooling and physical security. The network equipment us typically from the data centre, but part of it can also be from the customer.

- Bare metal
  With a bare metal service the customer rents a physical server that is under total control of the customer. The data centre operator cannot access or influence in any way the software running on that server.

- Virtual compute
  The last a largest group is that of the virtual compute. These are digital services that cover offer all sorts of virtualization. Meaning the physical hardware is under control of the data centre. The data centre runs on that server a piece of software (often called platform) that offers virtual computations. Very common are virtual machines (VMs), but also other types as (Kubernetes) containers, apps on AppEngines, and many more are possible.

4.5 Data centre placement

From a sustainability perspective data centres have a very unique property that they can, in principle, be built at any location. That makes that sustainability aspects can also play a role in determining the location of a new data centre. Data centres can for instance be located next to a windmill park or a PV park to have actually access to green electricity when the wind blows or the sun shines at the specific location. Or next a residential area to re-use the residual heat from the data centre for heating of houses and buildings.

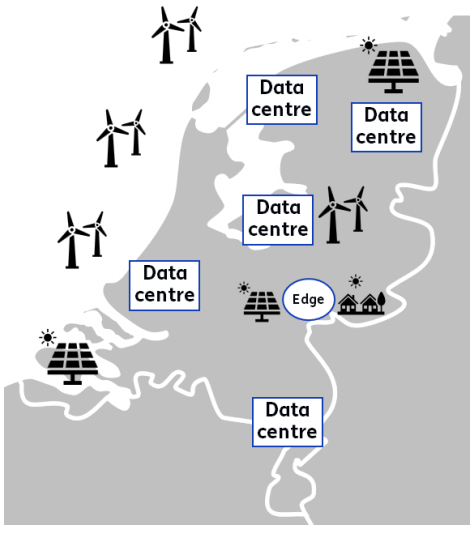

Figure 5: Example of data centres in relation to sustainable energy locations

5 Electricity system domain

This chapter focuses on the effects of the energy transition on both the electricity grid and on electricity markets and how that asks for more adaptable and flexible behaviour from entities connected to the grid.

5.1 Energy transition

The energy transition is targeted at greatly reducing the dependence on fossil fuels, and ultimately aims for carbon neutrality. The two main pillars that the energy transition relies on are electrification and renewable production.

Electrification is about replacing technologies or processes that originally relied on fossil fuels with electrical power. Examples of such technologies are transport (replacing combustion engines by electrical ones), heating (replacing gas boilers by full electric heat pumps) or the production of hydrogen (replacing gas by electrolysis with green electricity).

Electrification in itself does not necessarily reduce CO2 emissions if the electricity is still produced by conventional fossil power plants. For this conventional power plants will have to be replaced by renewable (and often distributed) production units.

Electrification and an ever growing share of renewable production fundamentally change the way the electricity grid has to be run. Two main challenges will have to be addressed:

- **Congestion.** Electrification puts a lot of additional load on the electricity grid for which it was not dimensioned. All this additional power cannot always be transported over the grid, leading to congestion. Another contributing factor to congestion is the intermittent and distributed character of renewable production. Solar and wind farms are located throughout the grid and can cause feed-in congestion on sunny and/or windy days.
- **Imbalance.** In the electricity grid production and consumption always have to be balanced to keep the grid frequency at 50 Hz. Traditionally, electricity production by power plants was scheduled to match the forecasted electricity consumption as closely as possible. Despite these forecasts, there are always differences between the actual consumption and production during the operational phase. These differences are bridged by keeping some capacity of power plants in reserve to react when needed. As already mentioned the production of renewable electricity typically has an intermittent character and cannot be controlled like traditional power plants. This makes it a lot harder to maintain the balance in the grid. There are fast reacting power plants that can provide balancing power, but they are operated on natural gas and cause emissions. Ultimately, this means that consumption of electricity should be much more flexible as it will have to adapt to the momentarily renewable production.

5.2 Congestion

In the Netherlands net congestion is becoming more and more of an issue. One of the reasons is that the Netherlands historically heavily relied on natural gas, e.g. for heating. This led to the roll out of electricity grids with a smaller capacity than in surrounding countries.

Figure 6 shows the capacity map of the Netherlands for the medium and high voltage grid. The colors represent the availability of transport capacity for consumption. In the red areas there is a shortage of transport capacity and companies that request a connection are put on a waiting list. New data centres, for example, will not be connected to the grid until the capacity of the grid is expanded. The expansion of a grid segment can take many years.

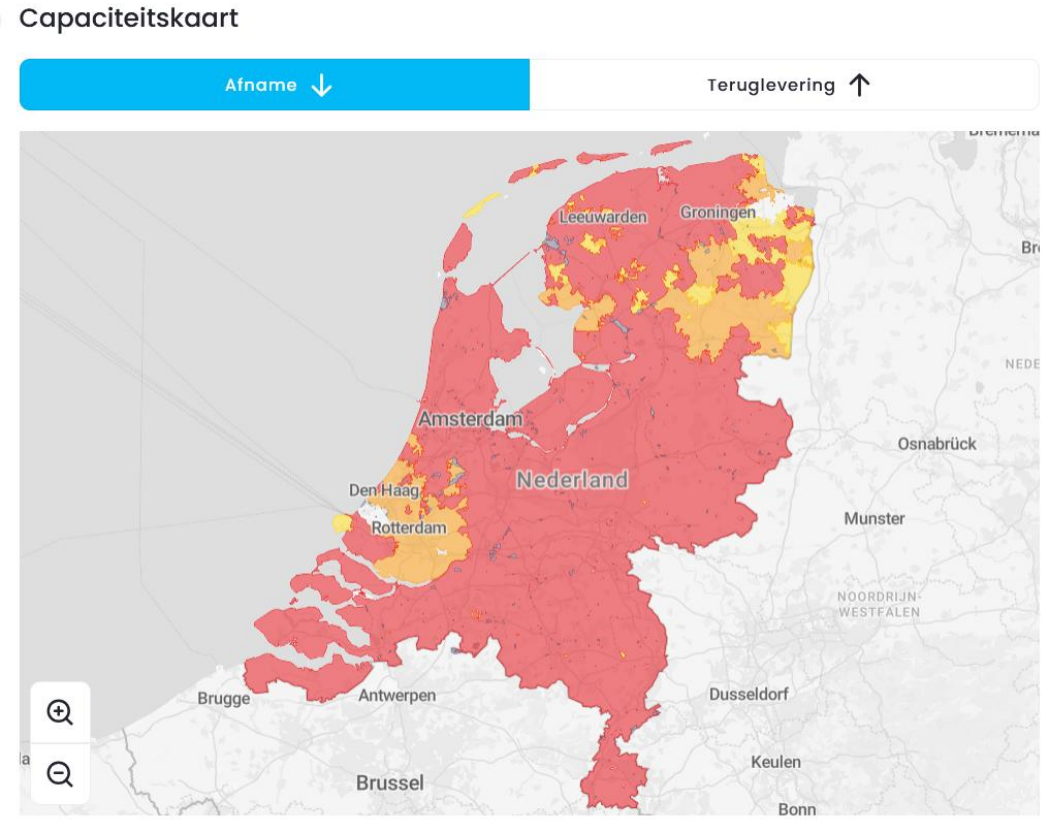

Figure 6: Capacity map of the Netherlands showing consumption congestion on the MV and HV grids[12]

In order to alleviate some of the consequences of congestion, DSO's have developed new contract forms whereby companies are no longer guaranteed the full capacity of their connection all the time. There are also contracts that arrange for the sharing a single connection between multiple companies; together they will have to make sure that do not exceed the capacity of the shared connection. In order to meet the requirements of these new transport contracts companies will have to build in some flexibility in the form of demand response or storage.

[12] Source: https://data.partnersinenergie.nl/capaciteitskaart/totaal/afname, visited on June 16th 2025

5.3 Energy markets

The relation with the DSO is part of a larger ecosystem that covers multiple stakeholders and a wide range of energy products and services. Figure 7 shows the most important stakeholders from a consumer (data centre) point of view.

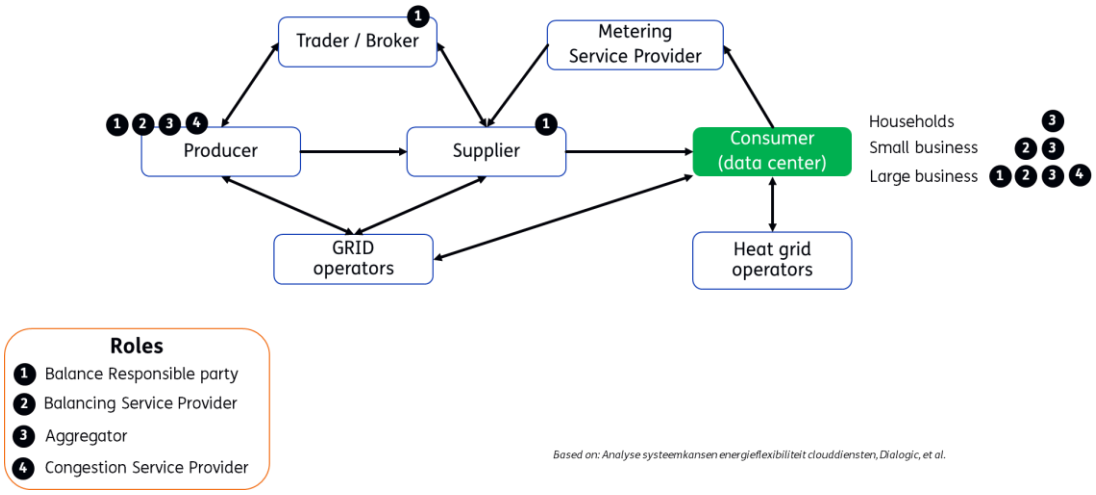

Figure 7: A simplified overview of relevant energy market stakeholders for a data centre

The picture also introduces 4 roles in the energy market which can be taken on by the different stakeholders. First these roles will be explained before moving on the description of the stakeholders. The following roles are mentioned:

- **Balance Responsible Party (BRP).** A BRP has to predict how much energy it will consume from the grid and/or how much energy it will feed in (produce) to the grid. This information must be submitted to the Transmission System Operator (TSO) on forehand so that it can check that the grid will be balanced. A BRP may be fined by the Transmission System Operator if turns out that it failed to meet these predictions during the operational phase.
- **Balancing Service Provider (BSP).** A BSP provides so called ancillary services to the TSO to make sure that grid is balanced at all times. BSP's should be able to quickly react to imbalances in the grid.
- **Aggregator.** An aggregator combines the flexibility of smaller producers/consumers into a much larger virtual unit that it can trade on various energy markets.
- **Congestion Service Provider (CSP).** A CSP provides services to Distribution System Operators (DSO's) and TSO's to help alleviate congestion. In practice this could mean that a CSP offers DSO's and/or TSO's the ability to temporarily lower its energy consumption at a certain price.

The following stakeholders are depicted:

- **Supplier.** The supplier delivers the energy to the data centre. There can be various contract forms, such as dynamic tariffs, which the data centre can choose from. The supplier also plays the BRP role. A supplier contracts producers (either directly or via the trader/broker) to produce the energy that it delivers to its customers.
- **GRID operators.** The Transmission System Operator (TSO) and the Distribution System Operator (DSO) are responsible for transporting the electricity from producers to

consumers over their grids. Where applicable, both the TSO and the DSO aim to alleviate congestions as much as possible. In order to do this they acquire congestion services from Congestion Service Providers. The TSO is also responsible for maintaining the balance on the grid and will ask Balancing Service Providers to provide so called ancillary services.

- **Trader/Broker.** This role facilitates the trading of energy volumes between producers and suppliers or also directly with large consumers. There are different energy markets such as the day ahead or the intraday market. Each market features its own set of energy products that can be traded.
- **Metering Service Provider.** This role is responsible for measuring the actual energy consumption and/or production by a consumer. This main application for this data is to be shared with the energy supplier for billing purposes.
- **Heat grid operators.** If a data centre is connected to a heat grid, this is the role that manages the heat grid.
- **Producer.** This role is responsible for producing the electricity and can be a conventional power plant or a renewable production unit. Dependent on its size it can be a Balance Responsible Party, a Balancing Service Provider and/or a Congestion Service Provider. The flexibility of smaller producers can be combined by an aggregator to be traded as one virtual production unit.
- **Consumer.** A data centre is a consumer in the energy market. Just like a producer it can take on the same roles, depending on its size. The flexibility of smaller consumers can also be combined by an aggregator to form a larger virtual consumer.

There is a wide variety of energy products and services that can be exchanged between the different energy stakeholders. All these products and services vary in the time scale on which they operate, the volumes and the expected behaviour from the stakeholders involved.

New products and services are constantly being developed to address the challenges of the energy transition. One such example is the GOPACS platform[13] that offers innovative congestion management services.

A comprehensive overview of balancing and congestion products and services and their characteristics can be found in the appendix 2 of the "Analyse systeemkansen energieflexibiliteit clouddiensten"[14] report by Dialogic, Pb7 and Entrance.

[13] https://www.gopacs.eu/en/
[14] https://topsectorenergie.nl/nl/nieuws/analyse-systeemkansen-energieflexibiliteit-clouddiensten/

6 Data Centre Flexibility Capabilities

This chapter investigates the capabilities that data centres have to provide energy flexibility. What mechanisms do data centres have to adapt their energy usage? What is the impact of each mechanism on the data centre and its customers?

Chapter 2 explained the different demands for flexibility: Congestion in the electricity grid, the unguaranteed availability of green electricity and the locality of green electricity. Chapter 2 also explains that the heat reuse customers do (in most cases) not request flexible heat because most heat reuse has a very constant heat demand with often only seasonal changes. Therefore heat is not treated as a demand for flexibility in this deliverable.

6.1 Data centre flexible electricity assets

In chapter 4 an overview is presented of the main assets of a data centre, split into compute and non-compute assets. Some of these can be used for electricity flexibility and others not or are not very likely. This section investigates these differences in more details and concludes with a list of non-compute and compute assets that can be used for electricity flexibility.

6.1.1 Flexible non-compute assets

A number of non-compute assets (from chapter 4) can contribute to the three reasons to offer electricity flexibility (from chapter 2):

Congestion in the electricity grid
There are a number of data centres assets that can produce electricity: PV panels, (diesel) generators and charged batteries[15]. Electricity production can be used to participate in the electricity grid congestion challenge. This is steered by the electricity grid often via financial benefits.

Generators are often fueled with diesel, which makes them not green but they can act as flexibility assets. Some data centres have their own PV panels installed. Compared to the energy consumption of entire data centre, PV panels contribute very little. For this reason they are not treated as a flexibility asset. Note that large solar farms next to a data centre can be seen as a flexibility asset, but those are typically not owned by the data centre itself.

Batteries can be used to make money via the electricity grid. Charge when the prices are low or even negative meaning the data centre even gets paid to charge the battery. Discharge when there is a shortage of electricity and the data centre gets a high(er) price for the delivered electricity.

[15] Charged batteries actually do not produce electricity but provide electricity by discharging.

Availability of green electricity & locality of green electricity
Additional batteries can be installed to be charged when there is a surplus of green electricity and when there is insufficient green electricity available on the grid be discharged to feed green electricity into the data centre instead of using it from the grid at that moment. In this way batteries can be used to provide green electricity for a longer period of time to the data centre. This can be for the data centre as whole, reducing to CO2-emissions for all customers a bit. Another option is to use the discharged green electricity only for a specific group of servers, sold as green services, reducing the CO2-emissions for a specific set of customers significantly.

Note that batteries used in this manner only applies for additional batteries. Most data centres do already have batteries installed but for the UPS functionality to bridge short (<15 min) electricity gaps. During longer gaps the (diesel) generators are started.

Summarizing
The following non-compute assets are candidates to be used for electricity flexibility:
- (Diesel) generators
- Additional batteries

6.1.2 Flexible compute assets

Next to the non-compute chapter 4 also describes a number of compute assets. This section investigates their possible contribution to the three reasons for electricity flexibility.

Compute contains assets for storage, networking and processing. For electricity flexibility controlled usage of the assets is needed, such a change actually influences the electricity consumption. The quicker and the more often an asset can react, the more value it has for flexibility purposes. Several times a day within minutes is realistic (see chapter 5 for more information).

This responsiveness requirement can (in principle) be delivered by the computational assets by adapting the computational workloads running on these assets and thereby increasing or decreasing their electricity consumption. Using storage or networking assets for electricity flexibility is much less relevant.

Computational assets can be both physical assets: the bare metal servers, as well as virtual computational assets: virtual machines, containers, app-engines, etc. From a data centre perspective it sells computational assets for their customers to run software on. Each data centre can decide which type of virtual computational assets they want to deliver. The only assumption relevant for this deliverable is that this virtual asset can run computational workloads.

To adapt computational workloads for electricity flexibility there are three methods relevant:

- *Time shifting*
  By starting a workload at a later moment in time electricity usage can be pushed to a more suitable moment. Postponing or bringing forward computational workloads will influence the electricity usage pattern over time.

  When computational workload is shifted in time it also shifts the electricity consumption on a physical compute, e.g. a server. When the server has no workload at all, going to sleep or even turning it temporarily off saves even more electricity.

Note that not all computational tasks can be shifted in time. Continuous workloads have by definition no start and end time and can therefore not easily be shifted in time.

- *Speed scaling*
  By lowering the speed of computation the average (and in some cases also the overall) power consumption can be reduced.[16]

  Depending on the type of workload, the speed of computation can be influenced via the clock speed of the CPU and/or the number of cores assigned to the computational workload. From a higher software perspective the number of computational units (e.g. number of workers) can also be adapted. Each individual having the clock speed and/or number of cores option.

- *Location shifting*
  By moving the computational workload towards physical locations with a "better" electricity profile (no congestion and/or green electricity availability) electricity usage can be influenced in a desirable direction. Changing the physical location of the workload means for instance changing the data centre or moving to another edge location of the same data centre.

These three computational flexibility methods can be mapped on the three reasons for electricity flexibility (from chapter 2):

Congestion in the electricity grid
When there is congestion on the electricity grid the demand or production of electricity exceeds the capacity of the grid. Meaning a data centre should be able to sometimes reduce their electricity consumption and on other moments expand their electricity consumption to consume to temporal overproduction.

All three types of computational asset flexibility can be used for this:
- Time shifting – By delaying workload the electricity usage can be postponed.
- Speed scaling – By reducing the calculation speed the average electricity usage can be decreased.
- Location shifting – Computation can be moved to a physical location without congestion issues. Thereby reducing the congestion issues at the originating location.

Availability of green electricity
At the moments green electricity is not available, running fully green compute becomes impossible. From the compute perspectives there are only two options:
1. run the computational workload at a later moment in time when there is green electricity available. This works for both bare metal as virtual compute services.
2. move the virtual computational workload to another location where there is green electricity available and the correct compute environment is available to run the workload on. This does not work for bare metal because then workload cannot be moved easily.

Speed scaling by reducing the calculation speed is not op option for fully green compute because that still consumes electricity. For best effort green compute it might be an option.

[16] https://dl.acm.org/doi/10.1145/3679240.3735103

Speed scaling by increasing the speed when there is a surplus of green electricity could be a viable option.

Computational asset flexibility that can be used:
- Time shifting – Delay the workload until green electricity becomes available.
- Location shifting – Move the virtual workload to a location with green electricity and the correct compute environment.
- Speed scaling – Only upward speed scaling is useful for flexibility: increasing the speed when there is a surplus of green electricity.

Locality of green electricity
When green electricity is not available at a specific location there remains only one option: move the virtual computational workload to another location where there is green electricity available and the correct compute environment is available to run the workload on. Also in this situation the move does not work for bare metal because moving a physical machine from one location to another takes a lot time and is therefore not treated as flexibility.

Computational asset flexibility that can be used:
- Location shifting – Move the virtual workload to a location with green electricity and the correct compute environment.

Summarizing
The following compute assets are candidates to be used for electricity flexibility:
- All computational assets

6.2 Customer flexibility capabilities

From a flexibility perspective the customer has (at least) three capabilities to work with. The first is choosing which DC centre or centre is wants to use. Each data centre will have different services with different capabilities and options. When a customer has contracts with multiple data centres it can divide its compute load across these data centres and can therefore be part of a flexibility regime. So next to the data centre also the customer could have the capability of location shifting.

The second is building (or restructuring) the software to become suitable for flexibility. Software that is capable to deal with time shifting and speed scaling. Meaning parts of the software can run at a later moment and parts of the software can handle longer processing. Adapting the software is outside the scope of this deliverable, but essential when data centres start involving customers with the flexibility challenge. time. This is a topic for further research.

A third flexibility capability customers might have is providing flexibility via quality degradation of their services, that are running in the data centre. The idea is that a service is working at a lower quality but still provide functionality. For instance video streaming but at a lower resolution, Google search without the first AI based result. An useful paper on this topic is “Carbon-Aware Quality Adaptation for Energy-Intensive Services”.[17]

Customer location shifting is inside the scope of this deliverable mainly to inform that this type of location shifting is also possible and must not be confused with the capability of the data centre to do location shifting.

[17] https://dl.acm.org/doi/pdf/10.1145/3679240.3734614

6.3 Data centre flexibility from a customer perspective

Next to the capabilities of the data centre to provide electricity flexibility it is also important to look at the perspective of the customers of the data centre. Perhaps at some point it becomes necessary to involve them too in the flexibility challenge. For that it is important to understand what their motivations and requirements are.

First of all no customer wants to be forced to participate in flexibility. That will disturb their current way of working, costs probably more money and brings them nothing. A customer will only participate in flexibility if it serves him a purpose. Traditionally almost all customers requires *100% service availability*[18] at the *lowest possible price*. That provides them a stable ground to build their software and business on. That is what they outsource to a data centre.

Nowadays and looking at the expected legislation (see chapter 3) an additional customer requirement emerges: *low CO2 emissions*. Companies and organisations are getting more and more pressure to become more sustainable. A typical indicator to express sustainability is the CO2 emissions[19]. For those companies whose CO2 emissions come to a significant extent from IT, it becomes relevant that their data centres provide insight and influence. So next to 100% availability and low prices also low CO2 emissions becomes relevant for customers.

Unfortunately all three at the same time is not possible because green electricity (zero CO2 emissions) is not continuous available. All the techniques to address this cost money, making the services more expensive. Some techniques even will reduce the 100% availability. For instance a customer that only wants compute services with zero CO2 emissions must accept that such a service will not always be available, due to the not continuous availability of green electricity .

Summarizing
Three key requirements of data centre customers are: 100% availability, low prices and low CO2 emissions. Because low CO2 emissions are likely to become mandatory, flexibility will most probably also impact (in the near future) the customer in some form.

Depending on the customers priority between these three requirements, different compute service offerings are required. Chapter 7 will go into more details about possible service offerings for sustainable flexible compute.

[18] 100% availability is not possible but data centres strive to high availability numbers with multiple 9 behind the comma. For simplicity this deliverable uses the term 100% availability.

[19] Note in this deliverable we exclude nuclear power plants although they do not produce any CO2 emissions. Most countries do not have such power plants and/or enough to provide continuous green electricity for the entire national electricity demand. Beside that there is the issue of nuclear waste and if that makes nuclear power plants sustainable or not.

6.4 Combined view on data centre flexibility capabilities

This section provides a combined overview of the flexibility forces on a data centre and the flexibility capabilities of both the data centre and its customers. It combines the knowledge from chapters 2 and 3 what explained why flexibility is needed and the knowledge of this chapter on the flexibility capabilities of data centres and its customers. Figure 8 shows the these flexibility forces on a data centre.

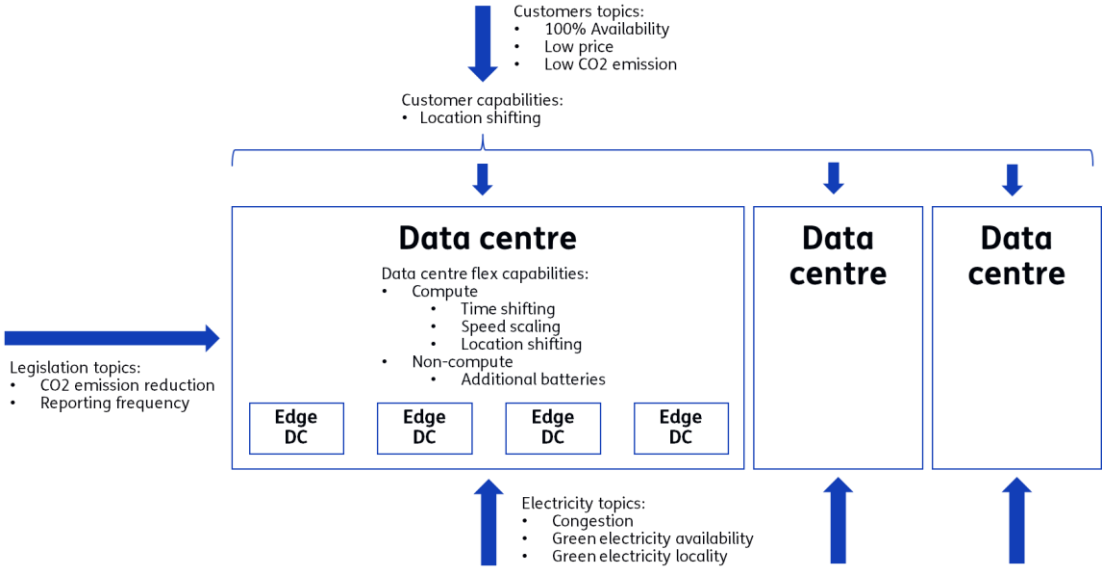

Figure 8: overview of flexibility forces on a data centre

From a customer perspective it can have contracts with multiple data centre providers. Location shifting helps him to decide which part of its software runs at which location. Decisions can be based on physical location, prices, availability of special resources (e.g. GPUs), etc.
A data centre provider might own multiple locations, in this picture called edge data centres. The data centre provider can internally use location shifting across his locations to move customer's compute loads to the best location at a moment in time.

At the bottom the electricity topics forcing the data centre to (re)think about its green electricity consumption. From the left the legislation topics that most likely will be enforced. And from the top the overall customer topics, divided over multiple data centres via location shifting.

Within the data centre it's the flexibility capabilities are shown, indicating the freedom the data centre has to make its operation green.

6.5 Towards a sustainable and flexible data centre

What would be a good first step towards more sustainable data centres? Although this question will have a different answer for each data centre, at a more generic scale at least three steps can be identified (not necessary in this order):

Step 1: provide insight
A very important step is to get insight in the sustainability of the data centre and in the context of this deliverable in the electricity consumption and CO2 emissions. At least for a data centre as a whole, but preferable at more detailed levels e.g. per customer, per server and ideally per computational workload. This is the finest grain a data centre has visibility over. Even more fine grained, up to the end user level, is out of scope for the data centre and must be calculated by the customer. Further research is needed to determine if and how a customer can do this and which information she needs (from the data centre) for those calculations.

Additionally besides insights on live data about electricity consumption, price and CO2 emission, also predictions would be very useful about the customer compute demand and the energy mix. This paves the way for orchestrating the non-compute and compute capabilities of the data centre in more predictable manner.

Step 2: Optimize without impacting the customer
Having insight in what happens and what might happen in the near future, it becomes possible to start using the data centre flexibility capabilities to influence the sustainability of the compute services provided to the customer. For instance by optimizing the internal availability of green electricity through additional batteries and assign that green electricity to the green compute services with priority. The customers will notice this immediately via insights provided to them according to step 1.

Perhaps also the compute location shifting capability of the data centre can be used for this. Moving compute load to locations with a surplus of green electricity. But doing this without impacting the customer is not trivial. It will depend on the actual SLAs about downtime and hick ups if it is feasible. Time shifting and speed scaling are not possible at this stage because they influence the customer directly as they temporally have no compute or slow compute.

Note that in a certain way some data centres are already working on this step by trading on the imbalance and congestion electricity markets and making money with it.

Step 3: Optimize with customer involvement
In this step also the customer will be involved / impacted by the sustainability measures. This can be achieved in a basic and a more advanced manner:

Basic
The data centre informs the customer upfront on the availability and speed of green compute services. The customer can decide if it wants to make use of it or not.

Advanced
In this variant there is a continuous “negotiation” between the customer and the data centre.

- The customer provides insight in his flexibility possibilities, their flex space, to the data centre. Meaning for instance they provide insight in when they need compute and when they can wait. Or even more advanced they inform the data centre about the latest moment his compute task must be finished and some indication on how heavy the compute task is. This is still very difficult to perform, but would be a very nice worry-free service towards the customer.
- The data centre combines all flex spaces of all customers and all other relevant data
- Then the data centre chooses the best point in the flex space of each customer. This way the data centre can make compute service as green as possible, each data centre using its own decisions and algorithms to way different interests.
- Depending on the agreement, the customer is then "asked" or "forced" to keep to the chosen point in the flex space.

All data centre's flexibility capabilities - time shifting, speed scaling and location shifting – are useable for this. The next chapter will investigate the different step 3 approaches into more detail.

7 Flexibility interfaces

This chapter discusses the different interfaces needed to incorporate electricity flexibility in the data centre and how to involve its customers.

7.1 Interfaces between the data centre and electricity grid

This section explains the different interfaces needed between the data centre and the electricity grid to implement the electricity flexibility as described in this deliverable.

Electricity (already available)
A data centre needs several of interfaces with the electricity grid to be able to operate. The basic one that all electricity customers need whether they want to become flexible or not are a connection to power grid, to get the electricity in the building.

Pricing and congestion (already available)
Next pricing information is important. This could be a contract for a longer period of time or an API with more dynamic pricing information. Current electricity trading mechanisms are already suitable for dynamic pricing and congestion flexibility trading. See chapter 5 for more details.

With these interfaces each data centre can fulfill its electricity needs and participate in the congestion flexibility market.

CO2 emissions (already available)
Knowing if the electricity coming into the data centre at this moment in time is green or not, is important for green data centres. In order to determine that the current power mix is needed, indicating which electricity sources are used at this moment (last 15 min). Roughly they can be divided into renewable energy sources and conventional energy sources. It is the ratio of the renewable versus the conventional energy sources that determines the greenness of the electricity. This is expressed in the amount of emitted CO2. Ideally zero CO2 is emitted, indicating perfect green electricity.

Because the electricity grid is a full connected grid with producers and consumers all connected, the general approach is to consider the national level of the energy mix. For the Netherlands (and a number of other European countries) this information is collected and distributed by the "nationaal energie dashboard"[20], the national energy dashboard. This information is also provided via an API[21]. The important parameter is the "gr CO2/kWh". Expressing the gr of CO2 emitted per kWh in the current (real time) energy mix. This API also provides historical and prediction information. Historical on a hour, day, month and yearly basis. Prediction one day ahead.

[20] Totale elektriciteitsproductie | Nationaal Energie Dashboard
[21] Eigen toepassingen koppelen (API) | Nationaal Energie Dashboard

CO2 emission hourly certificates (not yet available)
What still needs to be developed are mechanisms to trade hourly CO2 emission certificates, called time based energy attribute certificates (T-EACs). They should have a timestamp of one hour and must be validated with meter, grid and CO2 emission data.

When T-EAC becomes available, it becomes possible to create a nation (Dutch) wide market to trade hourly electricity production certificates. With these certificates electricity consumers, e.g. data centres, can prove their actual emitted CO2 of their electricity usage.

Another possibility would be that electricity consumers get a target on how much CO2 they are allowed to emit. It is then up to the consumer (data centre) to determine how they stay below this target. For proving this the T-EAC are also needed.

7.2 Interfaces between the customer and the data centre

The data centre should provide an interface that focusses the customer key topics and the data centre flexibility capabilities. It should provide an interface to get compute services with 100% availability, low price and low CO2 emissions, based on time shifting, speed scaling and (data centre) location shifting. As stated in the previous chapter it is not possible to fulfill all three customer topics at the same, it will always be a limited combination. Depending on their priorities the customer must make a balanced choice.

Therefore the core of the customer facing interface must consist out of different compute service offerings, each being operated from a combination of 100% availability, low price and low CO2 emissions. Each of these compute service offerings should be implementable based on a combination of time shifting, speed scaling and (data centre) location shifting.

Designing good flexible compute propositions is something that must be learned over time and might differ from data centre to data centre. This deliverable provides a first set of virtual machine (VM) compute offerings to start experimenting with. Later also other compute service types like bare metal, containers, etc. must be added.

Figure 9 visualizes the combination of all flexibility interfaces. From the power grid to the data centre, showing its compute and non-compute capabilities. The first set of flexible service compute offerings all mapped onto the three customer topics. Each flexible service compute offering is described in more detail below this figure.

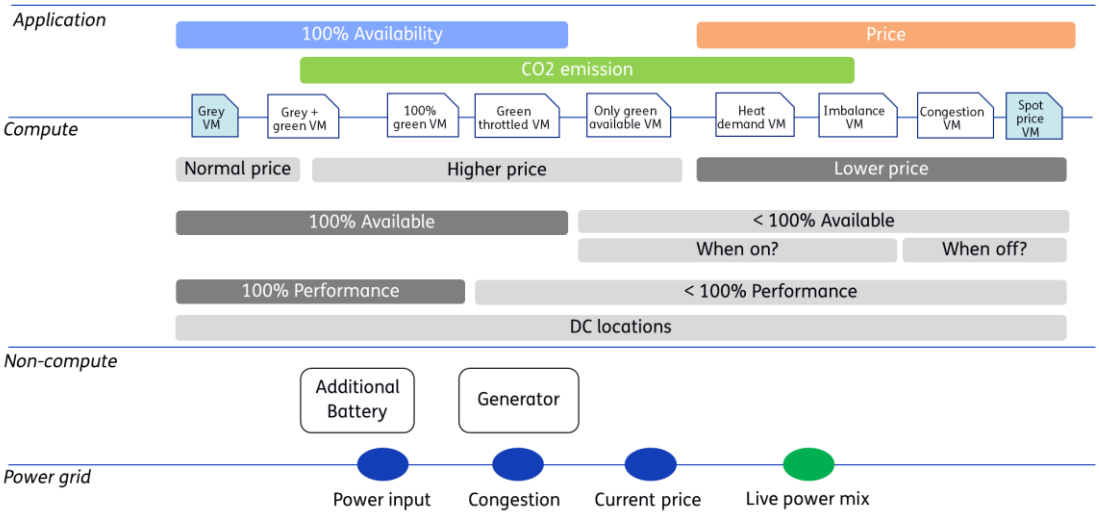

Figure 9: Flexible VM compute service descriptions

For reference two already in the market existing service compute offerings are shown:

- Grey VM
  The most "traditional" VM where it is unknown what the energy mix was. 100% available and against a reasonable price. In the context of this list it is called grey VM.

- Spot price VM
  A very cheap VM that is not always available. It is typically used by data centres to fill in unsold capacity on their servers. At the moment an paying customer arrives these VM is killed to make room.

Suggestions for new VM service descriptions

- Green + grey VM
  A VM with information on its actual energy mix. It is up to the data centre how much effort is put into making it more green. Default is the energy mix of the electricity grid.

- 100% green VM
  This VM goes a step further and guarantees 100% green and still being also 100% available. This means the data centre has mitigations in place for the moments the energy mix is not 100% CO2 free. Typically batteries. It is up to the data centre how much battery capacity they reserve for these VMs in order to sell them as 100% available and 100% CO2 emission free.
  Note that from a business perspective this might not be a realistic compute offering. The "green + grey VM" or the "Only Green available VM" (see below) are the more economical ones.

- Green throttled VM
  This VM is based on the 100% green VM, but adds the possibility that the data centre reduces temporarily the speed of the compute (CPU clock speed). Reducing the speed saves energy and thereby the VM can keep running on green electricity, although at a lower speed. Another approach would be that the VM is normally running at less than maximum speed and the speed is only increased when there is enough green electricity available for this VM.

- Only green available VM
  This VM is guaranteed 100% CO2 emission free, but not 100% available. It will go into sleep / suspend / shutdown when there is insufficient green electricity available.

- Heat demand VM
  When the data centre has a customer for its heat, earning additional money this way, it could offer heat VMs at a special price. Note that this is not a flexible controlled VM.

- Imbalance & Congestion VM
  These VMs contribute to either the imbalance or congestion market. It is mentioned as an option, probably not a realistic one.
  Note that it might imply having to stop spot VMs in order to make resources available for these flexibility VMs

7.3 Steps towards flexible compute APIs

Underlining the flexible service compute offerings must be an API the customer can use to communicate with the data centre about its compute service wishes and to maintain a continuous communication about the electricity flexibility.

Such an API must at least deliver:

- Insight in the customer's energy usage.
  Both in KWh and in CO2 emissions, because both are relevant for the customer to determine its next green actions.
- Control of the flexibility.
  A very promising methodology for this is the one already being used in communication about the electricity flexibility of households. There is a standard developed for this called S2[22] which is a communication standard for energy flexibility in homes and buildings. In short it is based on flexibility space. The household owner provides information on how much electricity flexibility it can provide, called the flex space. From the electricity grid side all flex spaces from many households are collected and an optimal choice for each of them is made maintaining a healthy electricity grid. That choice is sent back to the households, assuming (not binding) that they will follow.

Both these APIs are still in need of further design and (software) development.

[22] S2 - S2Standard.org

8 Conclusions

The main questions posed by this report were:

1. Where are the opportunities for data centres with regard to energy flexibility?
2. Where should this flexibility come from?

Starting with the first question, the conclusion should be that it is inevitable that data centres will have to provide much more flexibility in the years to come. Firstly, this has to do with the limitations of the electricity grid in terms of congestion and balancing challenges. This creates the opportunity for data centres to provide congestion and/or balancing services. For new grid connections for data centres, the provision of these grid services will probably not be voluntary but mandatory.

Another important contributing factor can be found in (EU) policies to further reduce emissions. These policies target a 55% reduction of emissions in 2030 (compared to 1990) and zero emissions in 2050. This means that the share of grey electricity will go down, necessitating all consumers (including data centres) to become more flexible to be able to match renewable production. This process will be relatively slow, but consumers can already be (close to) 100% green by buying guarantees of origins (GO's) for production by renewable sources that cover their consumption. Currently, these GO's are valid for a year which makes it relatively easy for consumers to be green on paper. However, there is a movement towards matching consumption and GO's on an hourly basis that is also supported by big tech companies such as Google and Microsoft. Should this be adopted in policy, then data centres will have to become flexible at a much faster rate.

Where should this flexibility come from? Currently, data centres can provide sufficient flexibility by using non-compute assets such as batteries. As explained above, the demand for flexibility will increase significantly in the coming years making it inevitable that some of that flexibility will also have to be provided by compute assets. It is not hard to imagine that this in turn will impact the services that data centres provide to their customers (and their respective end customers). Ideally, it should be a joint effort by data centres and their customers to explore what this impact means in practice.

ICT, Strategy & Policy

Zernikelaan 14
9747 AA Groningen
www.tno.nl

TNO innovation for life