



university of
 groningen

faculty of science
 and engineering

Exploring the Factors of Sustainability
 and creating a Sustainability Dashboard
 for BT Global Services

Internship Report

Author

Richard Westerhof
r.s.westerhof.2@student.rug.nl

Date

July 8, 2021

Supervisor

Vasilios Andrikopoulos

External Supervisor

Christos Bantis

Secondary External Supervisor

Richard Atherton

Contents

1	Introduction	2
2	State of the Art	2
3	Methodology	3
3.1	Selecting Metrics	3
3.2	Planning	3
4	Implementation	3
5	Discussion	5
5.1	Self Reflection	5
5.2	Current Limitations	5
6	Conclusion	6
6.1	Summary of the Project	6
6.2	Future Work	6
	References	7
A	List of Metrics	9
A.1	Energy Efficiency	9
A.1.1	Power Usage Effectiveness	9
A.1.2	Data Center Infrastructure Efficiency	9
A.1.3	Compute Power Efficiency	10
A.1.4	Data Center Energy Productivity	10
A.1.5	Data Center Fixed to Variable Energy Ratio	10
A.1.6	Green Energy Coefficient	11
A.1.7	Data Center Performance Per Energy	11
A.1.8	Data Center Workload Power Efficiency	12
A.1.9	Server Compute Efficiency	12
A.1.10	Space, Watts and Performance (SWaP)	12
A.1.11	IT Usage Effectiveness	13
A.1.12	Total Power Usage Effectiveness	13
A.2	Cooling, Thermal and Air Management	13
A.2.1	Data Center Cooling System Efficiency	13
A.2.2	Grid Utilization Factor	14
A.2.3	Data Center Temperature	14
A.2.4	Airflow Efficiency	14
A.2.5	Return Temperature Index	15
A.2.6	Return Heat Index	15
A.3	Miscellaneous Metrics	15
A.3.1	Bits per Joule	15
A.3.2	Storage Efficiency	16

1 Introduction

BT Global Services is a Software as a Service (SaaS) provider that makes use of cloud service providers with data centers all around the globe.

Becoming carbon neutral is a big goal for most companies, including BT's customers. Part of achieving that goal is using energy efficient data centers for their company. However, most cloud service providers do not offer clear and extensive insights into their customers' usage numbers or carbon emissions. Thus the end-goal for BT is to have a dashboard that they can show to their customers, in order to help them first get an idea of their IT carbon footprint, and next to help them reduce that carbon footprint by showing areas in which energy is going to waste or is not being used efficiently.

This kind of information is valuable to BT as well, since they themselves are also a company that wants to lower its carbon emissions.

2 State of the Art

Environmental sustainability has already been researched many times, including studies that focus on data centers in particular [9, 12, 13, 17]. Because of this, there are many metrics documented which measure different factors of environmental sustainability. A list of metrics that were considered for this project can be found in [Appendix A](#). This list includes the name of the metric, the original work in which it was presented, the formula used to calculate it, a description of what the metric means, as well as potentially some of its limitations, and whether it is used on the dashboard or not, and, if it is not used on the dashboard, a short explanation why the metric is not used.

Gill and Buyya have performed research on sustainability metrics for data centers in [9]. In this research they collected and categorized a set of metrics, and proposed a conceptual model for sustainable cloud computing. This model is designed to be holistic, by covering application management, workload and resource management, infrastructure management, cooling management, energy consumption management, and thermal aware aspects.

Jagroep et al. have performed research into developing one metric that measures the entire resource utilization of a machine, called the Resource Utilization Score (RUS) [13]. This work mostly focuses on the resource utilization itself, and less on how it effects or describes the environmental sustainability aspect of it.

Jagroep et al. also performed a case study on the resource usage and energy consumption of two software products during their development, with the aim of actively integrating these factors into the regular software development process [12]. During this project, they also used a simple dashboard which showed the system's resource usage, such as the CPU, RAM, HDD, and network, as well as the energy consumption, and the Resource Utilization Score as documented in their earlier work from 2016 [13].

Reddy et al. have also performed research on metrics for sustainable data centers in [17]. They have collected and a vast set of metrics and categorized them into nine different categories. From their experience with this vast set of metrics they come to the conclusion that there is no single metric which perfectly describes every aspect of sustainability. They also suggest an automated platform could be built to collect data and calculate a set of metrics from this data. Essentially, creating this suggested platform is the goal of this internship project.

3 Methodology

3.1 Selecting Metrics

Since we are working with real world data centers that were not initially designed or set up to measure their environmental sustainability, some data simply could not be obtained in the time frame of this internship, as it would require extra measuring hardware to be installed. Therefore, selecting the metrics, to a large extent, came down to which ones we were able to get the data for.

Another important factor was whether the metrics are intuitive and easy to understand. Since the end goal for BT is to show this dashboard to their real customers, it should be easy to understand without needing too much explanation. Ideally, the metrics are either self explanatory, or can be explained in just one or two sentences.

All of the metrics that were considered for this project are listed in [Appendix A](#). This list also specifies which metrics did actually end up being used on the final dashboard, and which ones were not included along with the reason they were not included.

3.2 Planning

The planning for this internship project can be found in [Table 1](#). This planning was successfully followed with no major deviations.

Goal	Goal For	Deadline
Phase I: Research Phase		
Gather initial list of existing metrics.	RW	11 th of May, 2021
Phase II: Design Phase		
Check which metrics are measurable and which are not.	BT	18 th of May, 2021
Obtain first part of data center data.	BT	19 th of May, 2021
Finalize list of metrics to be used.	BT & RW	21 st of May, 2021
Finalize design mock-up of the dashboard.	RW	24 th of May, 2021
Phase III: Implementation Phase		
Obtain final part of data center data.	BT	31 st of May, 2021
Agree on the format of the CSV files.	BT & RW	7 th of June, 2021
Implement an initial version of the dashboard.	RW	21 st of June, 2021
Implement the tool to import CSV files.	RW	21 st of June, 2021
Finalize dashboard implementation and internship report.	RW	12 th of July, 2021

- RW stands for Richard Westerhof.
- BT stands for BT Global Services.

Table 1: The planning for this internship project.

4 Implementation

This project is set up using a few simple, yet powerful open source technologies. The technology stack is intentionally basic, since developing the dashboard is only part of this internship. The back end for the dashboard is built with Elasticsearch version 7.12.1¹. Elasticsearch was chosen because it natively integrates with Kibana, which is what was used for the dashboard’s UI. To ensure optimal compatibility, for Kibana the same version was used as the Elasticsearch version, version 7.12.1². Kibana was chosen because it is easy to create designs in, and has support for Vega visualizations, which in turn allow us to write scripted metric

¹<https://github.com/elastic/elasticsearch/tree/7.12>

²<https://github.com/elastic/kibana/tree/7.12>

aggregations³ to Elasticsearch. Essentially, this way we can fully program the calculations for the metrics into the dashboard UI itself. This is useful because it means that we can use Elasticsearch as our entire back end solution, rather than only using it as a database with a custom back end solution connecting it to Kibana. This also allows us to use the great native integration directly between Elasticsearch and Kiana. On top of that, it nicely separates the logic; Elasticsearch only provides the data it is asked to provide, while Kibana requests the exact data it needs to show the metrics on the dashboard UI. Elasticsearch and Kibana are deployed on BT's servers and are only accessible via BT's private network.

In order for BT's staff to import the data from their data centers into the dashboard's database, a tool had to be created which takes CSV files as input, and translates those to JSON. The reason data is initially provided in CSV format is because it is easy to create from a spreadsheet and also easier to open in a GUI application, which is what BT's staff is most comfortable with. Elasticsearch, however, requires data to be inserted through HTTP requests to a specific URL on its REST API, with a JSON body containing the data. Thus, the need for a tool to perform this conversion arose. Due to the process to gather and import the data still involving some manual labor, data is currently collected as an average per month, see [subsection 5.2](#) and [subsection 6.2](#).

The tool used to import CSV files into Elasticsearch is written in Kotlin, since it is a modern language that can be compiled into a `.jar` file with all dependencies included, meaning users only have to obtain and run the `.jar` file with no additional setup, which is highly user friendly.

A screenshot of the dashboard is provided in [Figure 1](#). Due to the dashboard containing sensitive data about BT's customers, fictional values have been used to produce this screenshot which do not represent any real data. This has the added effect that some values are quite unrealistic.

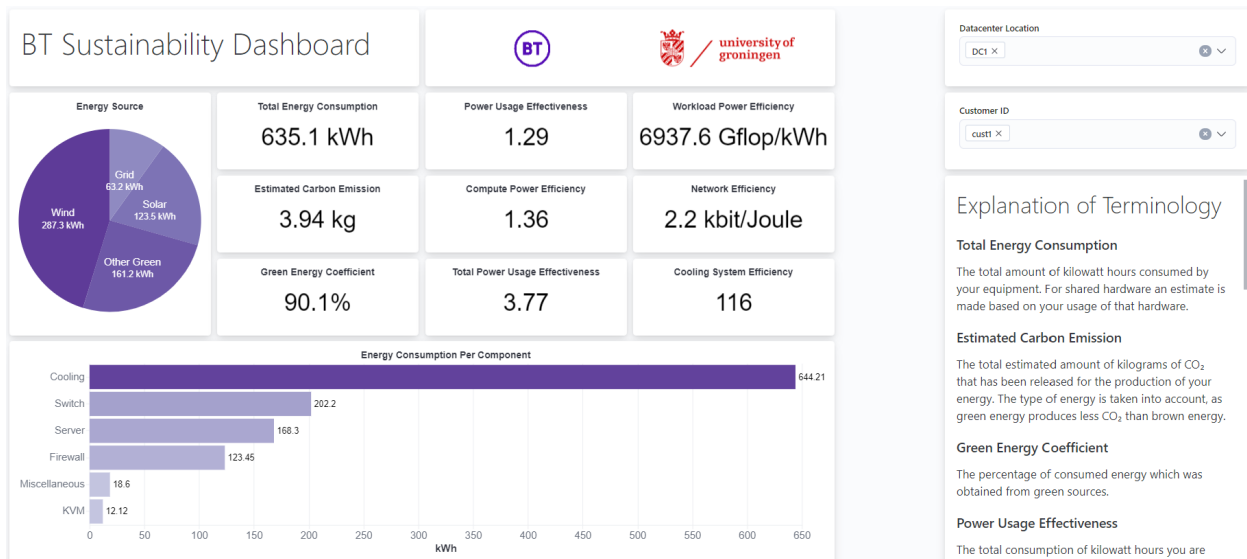


Figure 1: A screenshot of the dashboard's UI, using fictional data to protect BT's customers' privacy

The dashboard is horizontally separated into two sections. The left section contains nine metrics and two graphs. A description of each metric can be found in [Appendix A](#). The pie chart shows the distribution of energy consumption from four different kinds of energy, namely grid energy, solar energy, wind energy, and other, miscellaneous green energy sources. The bar chart shows the total energy consumption per device type.

The right section of the dashboard contains two filtering options in the form of multi-select dropdown

³<https://www.elastic.co/guide/en/elasticsearch/reference/7.12/search-aggregations-metrics-scripted-metric-aggregation.html>

menus. These allow the user to filter by data center location, since BT makes use of multiple data centers across the globe, and by customer ID, which should only contain customer IDs of customers of whom the current user is allowed to see the data. Essentially, this means that if the dashboard were to be opened to customers (see [subsection 6.2](#)), these customers should only be able to select their own data, but BT staff should be able to select any customer’s data. In [Figure 1](#), the data is filtered by data center location “DC1” and customer ID “cust1”. Below these dropdown menus, there is a list with short descriptions of each metric.

5 Discussion

5.1 Self Reflection

During this project I gained experience with working with a real company and real world data. This has improved my ability to develop a software product for a client, including communicating with them about the project, getting requirements, including them in the iteration process from the initial design to the final implementation, and turning their wishes into a product. I also gained experience with deploying the developed system in a real data center, which I had not done before.

Overall I think the internship went well; all of the goals we set were able to be completed, and the planning that was created at the start of the project was followed very closely. Miscommunications were rare, and when they did happen were resolved quickly. The scope of the project was also kept reasonable, as BT’s end-goal is quite ambitious. The dashboard that was created during this project serves mostly as a proof of concept and an initial foundation for future expansions.

5.2 Current Limitations

The main limitation of the dashboard is the data that the data centers can provide. Since their infrastructure was never designed to collect data regarding environmental sustainability, a lot of metrics were impossible to use due to the data not being available.

Another limitation is the current timescale of the data, which is one month. This means that every data point we have, represents an entire month’s usage. While this is fine for viewing total energy consumption or estimated carbon emissions, it does not allow for detecting patterns, at least, not on a scale smaller than multiple months. For example, it would be possible to see that the cooling energy consumption might be higher during the summer months, but it would not be possible to see whether resource usage is roughly constant during the day or includes high peaks during certain hours.

Furthermore, not all data that is collected is truly accurate. While the collected data is naturally gathered as accurately as possible, there are some values which are known to slightly differ from the real values. An example of a value which is higher than the real value is the cooling energy consumption, since there are a few other factors, for example the fire alarm, which are included in the value the data center provides for this metric, but are not separable⁴.

Additionally, data is currently only being collected from one of the data centers that BT uses. This means first of all that the data that is being shown is only part of the worldwide total that a customer may expect to see. Next, while theoretically collecting the same data from the other data centers should not be a problem, in reality these data centers are owned by different entities with possibly different measuring hardware or software, meaning that it remains to be seen whether the same data can actually be collected from these different data centers or not.

Because of all of these limitations, the dashboard is currently only suitable for internal use at BT. While the ultimate end-goal for BT with their sustainability project is to have a dashboard open to their customers,

⁴Though the consumption of the other devices should be negligible compared to the cooling energy consumption.

for this internship project it was determined early on that this would not be possible in the given amount of time.

6 Conclusion

6.1 Summary of the Project

In conclusion, the goal of this project was to explore the state of the art in data center environmental sustainability metrics, and then use those metrics to create an environmental sustainability dashboard for BT Global Services. This goal has been achieved by listing the state of the art metrics in [Appendix A](#) and building a dashboard using Elasticsearch and Kibana and deploying it on BT hardware. Furthermore, to make importing data from the data centers into the Elasticsearch database easier, a small program was created in Kotlin which reads the data from CSV files and makes the appropriate requests to Elasticsearch to insert the data and possibly update existing data, and was shared with BT. The source code for this tool will remain private.

6.2 Future Work

In the future, the dashboard could be extended to become a live dashboard, where the metrics are updated much more frequently, for example every minute. This would require a few changes, but they should be relatively simple ones. Currently, the data is given as an average figure for an entire month. However, some metrics require a concrete total, so the given average gets scaled up to a month before being inserted into the database. If the timescale were to change to, say, one minute instead of one month, the scaling of the averages would have to be changed as well. Though, this would make the new dashboard incompatible with the old, monthly data. Therefore, this would mean that either the old data has to be discarded, or that the old data has to be interpolated to minutely data, or that an extra field has to be added which specifies the timescale. The third option seems the most sensible, since it makes the dashboard compatible with any arbitrary timescale, however this would add more complexity in the calculations.

Another possible extension would be to allow BT's customers to view the dashboard themselves. This would require a few modifications to the current setup. The most important change would be to enable Elasticsearch's document level security⁵ to make sure that customers can only see documents that are attached to their own customer ID. This would prevent customers from being able to view other customers' data, which is obviously imperative. The reason why this was not implemented yet is because it requires a license for Elasticsearch, which was not obtained since the current version of the dashboard is only meant to be used as a proof of concept and only for internal use at BT. Furthermore, TLS should be enabled on the Kibana server, since sensitive information is being exchanged between the client and the server.

Naturally, further research could be done into finding more metrics to use on the dashboard. If this is done, care would need to be taken not to make the dashboard too complex, as currently it is rather simple and fits on one page with no to minimal scrolling required. Of course, new metrics can also replace some of the old metrics in case they are deemed more informative.

⁵<https://www.elastic.co/guide/en/elasticsearch/reference/7.12/document-level-security.html>

References

- [1] Green grid metrics: Describing data center power efficiency. *The Green Grid*, 2007.
- [2] Thermal guidelines for data processing environments – expanded data center classes and usage guidance. 2011.
- [3] New data center energy efficiency evaluation index DPPE (datacenter performance per energy) measurement guidelines. 2012.
- [4] A. I. Aravanis, A. Voulkidis, J. Salom, J. Townley, V. Georgiadou, A. Oleksiak, M. R. Porto, F. Roudet, and T. Zahariadis. Metrics for assessing flexibility and sustainability of next generation data centers. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6, 2015.
- [5] V. Avelar, D. Azevedo, and A. French. PUE: A comprehensive examination of the metric. *The Green Grid*, White Paper 49, 2012.
- [6] D. Azevedo, A. Hawkins, Z. Ortiz, R. Tipley, S. van den Berghe, and M. Blackburn. The Green Grid data center compute efficiency metric: DCcE. volume White Paper 34, 2010.
- [7] C. Belady and C. Malone. Metrics and an infrastructure model to evaluate data center efficiency. In *ASME 2007 InterPACK Conference collocated with the ASME/JSME 2007 Thermal Engineering Heat Transfer Summer Conference*, pages 751–755, 2007.
- [8] S. Cahal, K. Anandarao, C. Peters, S. Healy, N. Wayman, and S. Owen. Implementing cloud storage metrics to improve IT efficiency and capacity managements. 2011.
- [9] S. S. Gill and R. Buyya. A taxonomy and future directions for sustainable cloud computing: 360 degree view. In *ACM Computing Surveys*, volume 51, 2018.
- [10] D. Greenhill. Space watts and power. *EPA Power Efficiency Forum*, 2006.
- [11] M. K. Herrlin. Improved data center energy efficiency and thermal performance by advanced airflow analysis. In *Digital Power Forum*, 2007.
- [12] E. Jagroep, J. Broekman, J. M. E. M. van der Werf, S. Brinkkemper, P. Lago, L. Blom, and R. van Vliet. Awakening awareness on energy consumption in software engineering. In *IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Society Track*, pages 76–85, 2017.
- [13] E. Jagroep, J. M. E. M. van der Werf, J. Broekman, S. Brinkkemper, L. Blom, and R. van Vliet. A resource utilization score for software energy consumption. In *4th International Conference on ICT for Sustainability*, pages 19–28, 2016.
- [14] P. Mathew, S. Ganguly, S. Greenberg, and D. Sartor. Self-benchmarking guide for data centers: Metrics, benchmarks, actions. 01 2009.
- [15] L. Newcombe, Z. Limbuwala, P. Latham, and V. Smith. Data centre fixed to variable energy ratio metric DC-FVER. 2012.
- [16] M. K. Patterson, S. W. Poole, C.-H. Hsu, D. E. Maxwell, W. Tschudi, H. Coles, D. J. Martinez, and N. Bates. Tue, a new energy-efficiency metric applied at ornl’s jaguar. *Lecture Notes in Computer Science*, pages 372–382, 1 2013.
- [17] V. D. Reddy, B. Setz, G. S. V. R. K. Rao, G. R. Gangadharan, and M. Aiello. Metrics for sustainable data centers. In *IEEE Transactions on Sustainable Computing*, volume 2, pages 290–303, 2017.
- [18] V. Rodoplu and T. H. Meng. Bits-per-joule capacity of energy-limited wireless networks. In *IEEE Transactions on Wireless Communications*, volume 6, pages 857–865, 2007.

-
- [19] L. Sego, A. Márquez, A. Rawson, T. Cader, K. Fox, W. Gustafson, and C. Mundy. Implementing the data center energy productivity metric. *ACM Journal on Emerging Technologies in Computing Systems*, 8(4), 2012.
- [20] R. Sharma, C. Bash, and C. Patel. Dimensionless parameters for evaluation of thermal design and performance of large-scale data centers. *Proceedings of The IEEE - PIEEE*, 06 2002.
- [21] L. Siso, J. Salom, E. Oro, G. da Costa, and T. Zilio. CoolEmAll d5.6 final metrics and benchmarks. Technical report, IRIT-Institut de recherche en informatique de Toulouse, 2014.
- [22] T. Wilde, A. Auweter, M. K. Patterson, H. Shoukourian, H. Huber, A. Bode, D. Labrenz, and C. Cavazoni. DWPE, a new data center energy-efficiency metric bridging the gap between infrastructure and workload. In *International Conference on High Performance Computing & Simulation (HPCS)*, pages 893–901, 2014.

Appendix A List of Metrics

This appendix lists the metrics of the state of the art, with the source they were published in, the formula that calculates the metric, with a short description of each parameter, a description of the metric itself, and whether it is used on the current version of the dashboard or not. In case a metric is not currently used, a short justification for why it is not currently used is also provided. The metrics will be divided into multiple categories. These categories are: energy efficiency ([subsection A.1](#)), cooling, thermal and air management ([subsection A.2](#)), and miscellaneous metrics ([subsection A.3](#)), which includes a network metric and a storage metric.

A.1 Energy Efficiency

Energy efficiency includes metrics that describe the general energy efficiency of a data center. This category contains the largest number of metrics.

A.1.1 Power Usage Effectiveness

Source: [1, 4, 5, 14, 21]

Formula: $PUE = \frac{E_{DC}}{E_{IT}}$, where:

- PUE is the Power Usage Effectiveness.
- E_{DC} is the total energy consumption of the data center, in kilowatt hours.
- E_{IT} is the total energy consumption of the IT hardware, in kilowatt hours.

Description: Power Usage Effectiveness is one of the most commonly used metrics for data center environmental sustainability. It describes how much energy the data center consumes compared to the actual IT hardware⁶. Ideally, all energy in a data center would go to the IT hardware, thus the optimal value of this metric is one.

Used on dashboard: Yes.

A.1.2 Data Center Infrastructure Efficiency

Source: [1, 5, 21]

Formula: $DCIE = \frac{E_{DC}}{E_{non-IT}}$, where:

- $DCIE$ is the Data Center Infrastructure Effectiveness.
- E_{DC} is the total energy consumption of the data center, in kilowatt hours.
- E_{non-IT} is the total energy consumption of the non-IT hardware, in kilowatt hours.

Description: Data Center Infrastructure Efficiency is the complement to [Power Usage Effectiveness](#). It describes how much energy the data center consumes compared to the non-IT hardware. As mentioned for the Power Usage Effectiveness, all energy in a data center would ideally go to the IT hardware, thus the optimal value for this metric is infinity.

Used on dashboard: No, as it does not provide any new information compared to the Power Usage Effectiveness metric.

⁶The definition of “IT hardware” is up to the implementation; in this project we consider IT hardware to be server units, switches, routers and firewalls.

A.1.3 Compute Power Efficiency

Source: [7]

Formula: $CPE = \frac{U_{IT}}{PUE}$, where:

- CPE is the Compute Power Efficiency.
- U_{IT} is the the utilization of the IT hardware, as a percentage.
- PUE is the **Power Usage Effectiveness**.

Description: Compute Power Efficiency integrates resource utilization with **Power Usage Effectiveness**. The idea is that all hardware needs to be powered even if only a small percentage of it is currently in use. The downside is that this is not truly accurate, since most computing hardware, though not all, consumes less power when the utilization is low.

When this metric is measured on the level of an entire data center, the utilization of individual machines should be weighted according to the total resources available to the machine. For example, if machine A has 4 cores, but is only actively using 1⁷, then the utilization of this machine would be 25%. Now assume machine B has 8 cores⁸, but is actively using 4 of them, then the utilization of this machine would be 50%. If we were to take an unweighted average of the utilizations of these machines, it would be $\frac{25\%+50\%}{2} = 37.5\%$. However, if we take the weighted average, it would be $\frac{25\% \cdot 4 + 50\% \cdot 8}{4+8} = 41.7\%$.

Used on dashboard: Yes.

A.1.4 Data Center Energy Productivity

Source: [1, 19]

Formula: $DCEP = \frac{UW}{E_{DC}}$, where:

- $DCEP$ is the Data Center Energy Productivity, in kWh/UCU.
- UW is the the useful work performed by the system, in Useful Computation Units (UCU).
- E_{DC} is the amount of energy consumed by the entire data center, in kilowatt hours.

Description: Data Center Energy Productivity shows how many units of useful work the data center performs per kilowatt hour. However, the definition of which work is and is not useful is up to the implementation, as well as the exact definition of one Useful Computation Unit. In the original paper from 2007, it is defined as one discrete action performed by the system, “such as an email transaction, the execution of a query, the completion of a simulation, etc.” [1]. Each of these actions would have a weight attached to them, to make larger actions worth more UCU in the final equation.

However, all of this makes this metric only meaningful if full context is provided, and not comparable across data centers.

Used on dashboard: No, since we would first have to define one UCU in our context, and then also provide this information to every dashboard user. This makes the metric too complicated and not intuitive enough.

A.1.5 Data Center Fixed to Variable Energy Ratio

Source: [15]

Formula: $FVER = 1 + \frac{E_{fixed}}{E_{variable}}$, where:

⁷for ease of example, assume this core is constantly at 100% utilization while the inactive ones are constantly at 0%

⁸Again, for ease of example, assume these cores are the exact same as machine A’s cores. In the real calculation on the dashboard, each machine’s theoretical peak floating point operations per second (FLOPS) is used as a weight.

- *FVER* is the Data Center Fixed to Variable Energy Ratio.
- E_{fixed} is the power consumption of the IT hardware while idle, in Watt.
- $E_{variable}$ is the power consumption of the IT hardware while under heavy load minus the fixed energy, in Watt.

Description: Data Center Fixed to Variable Energy Ratio describes how much energy the IT hardware consumes while idle versus under heavy load. The goal is to minimize this metric, by reducing the energy consumption while idle.

Used on dashboard: No. Currently the data on the dashboard is collected as one number per month, which means we cannot find the minimum and maximum energy consumption rates for each month, unless they would have been explicitly provided as separate fields, but this was not done for this project.

A.1.6 Green Energy Coefficient

Source: [3]

Formula: $GEC = \frac{E_{renewable}}{E_{DC}}$, where:

- *GEC* is the Green Energy Coefficient.
- $E_{renewable}$ is the amount of renewable energy used, in kilowatt hours.
- E_{DC} is the total energy consumption of the data center, in kilowatt hours.

Description: The Green Energy Coefficient shows the percentage of the data center’s energy that is drawn from renewable energy sources.

Used on dashboard: Yes.

A.1.7 Data Center Performance Per Energy

Source: [3]

Formula: $DCPPE = ITEU \cdot ITEE \cdot \frac{1}{PUE} \cdot \frac{1}{1-GEC}$, where:

- *DCPPE* is the Data Center Performance Per Energy.
- *ITEU* is the IT equipment utilization, measured by taking the total measured energy consumption of all IT hardware and dividing by the rated energy consumption, in kilowatt hours.
- *ITEE* is the IT equipment energy efficiency, measured by taking the total amount of work done and dividing by the total rated power of the IT hardware. Here, work is defined by $\alpha \cdot \Sigma(\text{server performance [GTOPS]}) + \beta \cdot \Sigma(\text{storage capacity [GB]}) + \gamma \cdot \Sigma(\text{network capacity [GB/s]})$, and $\alpha = 7.72[W/GTOPS]$, $\beta = 0.0933[W/GB]$, $\gamma = 7.14[W/(GB/s)]$ ⁹
- *PUE* is the **Power Usage Effectiveness**.
- *GEC* is the **Green Energy Coefficient**.

Description: The Data Center Performance Per Energy metric is essentially a combination of four other metrics. By constructing this metric, the idea is that one value gives a broad overview of the data center’s environmental impact. It takes into consideration the IT equipment utilization, the server’s peak performance, storage capacity, and network speed, the **Power Usage Effectiveness** and the **Green Energy Coefficient**.

Used on dashboard: No, the values for α , β , and γ are effectively metrics of their own that can change over time, thus using the provided values is not an option. Furthermore, not all data to calculate these mini-metrics is available for this project.

⁹These values are based on “standard servers” from 2005 [3].

A.1.8 Data Center Workload Power Efficiency

Source: [22]

Formula: $DCWPE = \frac{FLOP}{E_{DC}}$, where:

- $DCWPE$ is the Data Center Workload Power Efficiency.
- $FLOP$ is the total amount of floating point operations performed.
- E_{DC} is the total energy consumed by the data center, in kilowatt hours.

Description: Data Center Workload Power Efficiency describes how efficiently the power consumption of an entire data center is used, according to the amount of floating point operations a data center performs per kilowatt hour of energy consumed.

Used on dashboard: Yes.

A.1.9 Server Compute Efficiency

Source: [6]

Formula: $SCE = T_{primary}/s$, where:

- SCE is the Server Compute Efficiency.
- $T_{primary}$ is the amount of time spent performing the server's primary task.

Description: Server Compute Efficiency describes how much of a server's time is spent performing its primary task. Other tasks, such as performing backups or updates, are considered secondary or tertiary tasks. However, while secondary and tertiary tasks may not be strictly necessary for the server to perform its primary task, they can still be essential to ensuring high quality of service (QoS). Furthermore, this is another metric that needs to be discussed in its own context, and cannot be used to compare any two arbitrary servers, since they may have different secondary and tertiary tasks that are used to improve the QoS.

Used on dashboard: No, similarly to the **Data Center Energy Productivity** metric, the definition of the server's primary task is up to the implementation, making this metric not intuitive.

A.1.10 Space, Watts and Performance (SWaP)

Source: [10]

Formula: $SWAP = \frac{FLOPS}{Space_{DC} \cdot E_{DC}}$, where:

- $SWAP$ is the Space, Watts and Performance rating of the data center.
- $FLOPS$ is the amount of floating point operations performed per second.
- $Space_{DC}$ is the amount of space the data center occupies, in cubic meters.
- E_{DC} is the power consumption of the data center, in Watts.

Description: Space, Watts, and Performance is a similar metric to the **Data Center Workload Power Efficiency**, but this metric also takes into consideration how much space the data center occupies, penalizing larger data centers if their performance does not scale with their size.

Used on dashboard: No, due to time constraints and delays in retrieving the needed data for this metric, this was not implemented on the dashboard. It would, however, be easy to add, as the data is currently being collected for it¹⁰.

¹⁰Though the formula for the metric would need to be slightly adjusted, as our data contains the floor area in square meters, rather than the volume in cubic meters

A.1.11 IT Usage Effectiveness

Source: [16]

Formula: $ITUE = \frac{E_{IT}}{E_{compute}}$, where:

- $ITUE$ is the IT Usage Effectiveness.
- E_{IT} is the power consumption of the IT hardware.
- $E_{compute}$ is the power consumption of the compute hardware.

Description: IT Usage Effectiveness is a metric similar to the **Power Usage Effectiveness**, but in a slightly different context. Whereas the Power Usage Effectiveness measures how much of the data center's power consumption is due to the IT hardware, the IT Usage Effectiveness measures how much of the IT hardware's power is due to actual compute hardware. The difference between IT hardware and compute hardware is that IT hardware includes networking equipment, storage devices, and other devices which are needed for the cloud service's operation. The compute hardware, however, is defined as only the components needed to perform computations, such as the CPU, GPU and RAM.

Used on dashboard: Yes, but indirectly. This metric is part of the bigger **Total Power Usage Effectiveness** metric, which is used on the dashboard. Therefore, this metric is not used by itself.

A.1.12 Total Power Usage Effectiveness

Source: [16]

Formula: $TUE = ITUE \dot{P}UE = \frac{E_{DC}}{E_{compute}}$, where:

- TUE is the Total Power Usage Effectiveness.
- $ITUE$ is the **IT Usage Effectiveness**.
- PUE is the **Power Usage Effectiveness**.
- E_{DC} is the amount of energy consumed by the data center, in kilowatt hours.
- $E_{compute}$ is the power consumption of the compute hardware, in kilowatt hours.

Description: Total Power Usage Effectiveness is an extension of the **Power Usage Effectiveness**, by adding the IT Usage Effectiveness to the equation. The goal of this metric is to measure how much power the compute hardware consumes compared to the power that the entire data center consumes. The compute hardware only includes hardware that is directly related to performing computations, such as the CPU, GPU (if present) and RAM.

Used on dashboard: Yes.

A.2 Cooling, Thermal and Air Management

The cooling, thermal and air management category mainly contains metrics related to temperatures and heat transfer.

A.2.1 Data Center Cooling System Efficiency

Source: [14]

Formula: $DCCSE = \frac{E_{cooling}}{CL_{DC}}$, where:

- $DCCSE$ is the Data Center Cooling System Efficiency, in kW/ton.
- $E_{cooling}$ is the power consumption of the cooling system, in kilowatt.
- CL_{DC} is the cooling load in the data center, in ton.

Description: Data Center Cooling System Efficiency, as the name suggests, shows the effectiveness of the cooling system. However, counter intuitively, this value should be as low as possible. When we look at the unit of this metric (kW/ton), this makes sense, as we want to minimize the amount of energy we have to spend on a given cooling load.

Used on dashboard: Yes.

A.2.2 Grid Utilization Factor

Source: [4]

Formula: $GUF = \frac{U_{grid}}{t}$, where:

- GUF is the Grid Utilization Factor, as a percentage.
- U_{grid} is the amount of time the energy grid was used, in seconds.
- t is the total amount of time that has passed, in seconds.

Description: The grid utilization factor is described by Aravanis et al [4] as the percentage of time that a data center consumes grid energy. This means that even if green energy sources can provide 99% of the system's energy consumption at a given moment, this still counts as time that the energy grid has to be used, thus increasing the grid utilization factor.

Alternatively, one could measure the amount of kilowatt hours consumed from grid energy and compare it to the total amount of kilowatt hours consumed by the data center. This is essentially the inverse of the **Green Energy Coefficient**.

Additionally, to obtain the 'Green Energy Utilization Factor', the percentage of time (or kilowatt hours, if the alternative formula is used) green energy is consumed, the formula is simply $GEUF = 1 - GUF$.

Used on dashboard: No, since we only collect on number for one month, and since this information is not measured directly, it would be impossible to calculate.

A.2.3 Data Center Temperature

Source: [2]

Formula: DCT , where:

- DCT is the Data Center Temperature, in °Celsius.

Description: The Data Center Temperature is not a metric that uses a formula, instead it can be directly measured. According to the original paper[2], the optimal range for the temperature that should be aimed for is between 18°C and 27°C. Keeping the ambient temperature between these values optimizes the lifespan and efficiency of the equipment.

Used on dashboard: No, the data for this metric was not available.

A.2.4 Airflow Efficiency

Source: [14]

Formula: $AfE = \frac{E_{fan}}{Af}$, where:

- AfE is the Airflow Efficiency, in kW/cfm.
- E_{fan} is the power consumption by the fans, in kilowatt.
- Af is the amount of airflow produced by the fans, in cubic foot per minute (cfm).

Description: Airflow Efficiency describes how efficiently the fans are able to push air around the data center. This depends on the own efficiency of the fans, but it also depends on the path of the air. Thus, by arranging equipment in a more airflow friendly manner, this metric can also be increased.

Used on dashboard: No, the data for this metric was not available.

A.2.5 Return Temperature Index

Source: [11]

Formula: $RTI = \frac{T_{return} - T_{supply}}{T_{exhaust} - T_{intake}}$, where:

- RTI is the Return Temperature Index.
- T_{return} is the temperature of the returned cooling air, in °Celsius.
- T_{supply} is the temperature of the supplied cooling air, in °Celsius.
- $T_{exhaust}$ is the temperature at the exhaust of the system, in °Celsius.
- T_{intake} is the temperature at the intake of the system, in °Celsius.

Description: The Return Temperature Index compares the difference between the temperature of the air that goes into the system and the temperature of the air that goes out of the system, to the difference between the temperature at the input of a rack and at the output of a rack. If the value is less than 100%, this indicates that some of the air is bypassing the racks, a value higher than 100% means hot air is recirculating.

Used on dashboard: No, the data for this metric was not available.

A.2.6 Return Heat Index

Source: [20]

Formula: $RHI = \frac{HE_{crac}}{ER_{exhaust}}$, where:

- RHI is the Return Heat Index.
- HE_{crac} is the heat extraction by the computer room air conditioning, in °C * m³/s.
- ER is the enthalpy rise at the air exhaust, in Joule.

Description: The Return Heat Index is in essence similar to the **Return Temperature Index**, it measures the thermal performance of a system. However, the Return Heat Index takes into consideration the airflow rate as well, whereas the Return Temperature Index focuses on temperature, like the name suggests.

Used on dashboard: No, the data for this metric was not available.

A.3 Miscellaneous Metrics

This category contains two miscellaneous metrics that did not fit in any of the other categories, but do not justify having their own separate categories.

A.3.1 Bits per Joule

Source: [18]

Formula: $BpJ = \frac{NT}{E_{network}}$, where:

- BpJ is the amount of bits that can be sent using one joule of energy, in b/J.
- NT is the amount of network traffic, in bits.
- $E_{network}$ is the amount of energy consumed by the network equipment, in Joule.

Description: Bits per Joule is a simple metric that shows how many bits are sent for every Joule of energy consumed by the network equipment.

Used on dashboard: Yes.

A.3.2 Storage Efficiency

Source: [8]

Formula: $SE = \frac{U_{storage}}{E_{storage}}$, where:

- SE is the Storage Efficiency, in Gigabyte/Watt.
- $U_{storage}$ is the storage usage, in Gigabytes.
- $E_{storage}$ is the amount of energy consumed by the storage media.

Description: Storage Efficiency shows how much data one Watt of power stores. Since storage media require the same amount of power to run regardless of how much data is stored on them, this metric can be optimized by having as little free space available as possible. Free space is considered to be wasted space that still draws power in this metric.

Used on dashboard: Yes.