

DC Energy Data Measurement and Analysis for Productivity and Waste Energy Assessment

1st Anastasiia Grishina
CNRS, CRAN, F-54000
University of Lorraine
Vandoeuvre-les-Nancy, France
anastasiia.grishina2@etu.univ-lorraine.fr

2nd Marta Chinnici
Energy Technologies Dep., ICT-Div.
ENEA-R.C. Casaccia
Rome, Italy
marta.chinnici@enea.it

3rd Davide De Chiara
Energy Technologies Dep., ICT-Div
ENEA-R.C. Portici
Portici (Na), Italy
davide.dechiara@enea.it

4th Guido Guarnieri
Energy Technologies Dep., ICT-Div
ENEA-R.C. Portici
Portici (Na), Italy
guido.guarnieri@enea.it

5th Ah-Lian Kor
School of Computing
Leeds Beckett University
Leeds, UK
A.Kor@leedsbeckett.ac.uk

6th Eric Rondeau
CNRS, CRAN, F-54000
University of Lorraine
Vandoeuvre-les-Nancy, France
eric.rondeau@univ-lorraine.fr

7th Jean-Philippe Georges
CNRS, CRAN, F-54000
University of Lorraine
Vandoeuvre-les-Nancy, France
jean-philippe.georges@univ-lorraine.fr

Abstract—The study and analysis of energy efficiency in Data Centers (DCs), through a set of globally accepted metrics, is an ongoing challenge. In particular, the area of productivity metrics is not completely explored, and there is no existing proposed metrics, which provides a direct measurement of the useful work in a DC. This paper proposes a methodology that addresses the problem of measurement, calculating, and evaluating the energy productivity assessment in Data Center (DC), which encompasses both the portion of energy employed for computing and energy wasted during computational work. It involves the estimation of productive energy consumption by a DC cluster based on the following: statistical data collection and interpretation, software for energy data analysis, and mathematical formulation. This current work is based on available data extracted through experiments conducted on the cluster “CRESCO4” from ENEA Data Center facilities. The dataset covers the power and job schedule characteristics running on the cluster for one year. This paper shows how to advance beyond state of the art for productivity metrics (e.g. useful work). It will also help enhance server performance and power management since the appropriate statistical data analysis provides a profile on server energy consumption behavior. Additionally, we make recommendations on how the productivity assessment could driver a new power efficiency management strategy, which is specifically targeted at DC manager and/or operators, and end-users of the facilities.

Keywords—Data Center, Energy Efficiency, Energy Consumption, Metrics, Cluster, Big Data, Data Analysis, Workload Management, Computer Application, Policies

I. INTRODUCTION

According to the Berkeley National Laboratory Report [1], data centers in the US consumed approximately 70 billion kWh electricity in the year 2014 (i.e. about 1.8% of total US

energy consumption). It is estimated that in the year 2020, the total energy consumption by the US Data Centers (DCs) will increase to 73 billion kWh. Initiatives such as the EU Code of Conduct for Energy Efficiency in DCs [2] have been established in response to the increasing energy consumption in DCs and to reduce their environmental impact. Best practice supplement to the Code of Conduct has been published in order to support DCs operators in their endeavor to improve the energy efficiency [3]. Monitoring of energy usage and consumption is essential to pursue the energy efficiency target and in the meantime to reduce the waste in the DC operations. However, metrics are required to provide insight into how efficient energy is transferred into IT equipment in a DC [4]. Additionally, metrics will enable DC operators to estimate their efficiency, draw a comparative analysis with others and also ascertain the need for further energy improvement [5]. PUE has become the de facto industrial “standard” metric for measuring DC energy efficiency because of its simplicity and ease of use. However, there are many shortcomings of the PUE metric related to the black box definition. Indeed, the PUE: would depict an accurate picture only if the IT equipment functions at its full capacity [6]; does not reveal the real energy performance of DCs, for e.g. IT equipment efficiency [7], [8]. Consequently, productivity metrics (e.g DCeP) have been created to address these gaps. This category includes indices related to the quantity of the “useful work” within a DC from an IT perspective. Hence, these indices should lead to questions such as: What is the “useful work” of a DC? and How does one calculate the “useful work” of a DC?

[9]. As an example: the amount of IT equipment necessary to meet a Service Level Agreement (SLA) within a specified response time [10]; IT equipment utilization (e.g. Compute Power Efficiency [11]; Server Compute Efficiency [12]).

This paper aims to demonstrate how productivity metrics could be applied in a real physical DC context. This will entail the investigation of productive energy consumption by the ENEA DC cluster. The following objectives will support this aim:

- a) Analyse the energy consumed to complete each process within the cluster during a specified period, T;
- b) Analyse the energy consumed by each aborted process within the cluster during the period, T;
- c) Analyse idle energy consumed within the cluster during the period T;
- d) Evaluate the load of the DC cluster based on (a–c) during the period T;
- e) Calculate energy efficiency and productive energy efficiency of the data centre cluster during the period T using the DCeP metric to calculate the useful work and the EWR metric for knowledge the waste energy (or “non-work”).

The organisation of this paper is as follows: Section I — Introduction; Section II — Related Work; Section III — Methodology; Section IV — Results and Discussion; Section V — Conclusions and Future Work.

II. RELATED WORK

The two primary factors for improving data center energy efficiency are the enhancement of facilities’ infrastructure efficiency and IT equipment/infrastructure efficiency (note: this is the focus of this paper). These efficiency measures can bring about savings concerning reduced electricity bill due to decreased operational expenditure (OpEx) and capital expenditure (CapEx) [13]. Diverse efforts have been made to reduce the energy consumption of IT equipment/infrastructure. Baccour and colleagues [14] have developed a new simulated green DC topology (known as PTNet) with power-aware routing algorithm. Actual server power consumption and utilization are measured to identify opportunities to increase DC compute capacity [15] and to determine the variation pattern of Power Usage Effectiveness (PUE) throughout an observed period [16]. Real-time monitoring for evaluation of DC energy efficiency has been conducted by [17], [18] through the use of comprehensive software tool for the acquisition and display of power consumption of associated IT equipment/infrastructure. Additionally, physical experiments have also been conducted to monitor the power consumption of servers followed by a quantitative analysis to estimate server power at the component level (i.e. CPU, memory, disk, etc.) [19].

Undeniably, power consumption monitoring is beneficial. However, the lack of appropriate metrics poses a significant barrier to improving energy efficiency in DCs [20]. Metrics will help communicate power consumption and also the environmental impact of data centers [21]. Several researchers have collated related environmental metrics for DCs (e.g. Carbon

Usage Effectiveness, CUE; Water Usage Effectiveness, WUE) [16], [22]. In order to improve the energy efficiency of DCs, it is necessary to look at metrics related to power, productivity, and server performance [20] and green metrics, which shall be further discussed below.

Power Usage Effectiveness (PUE) [4], [5], an example of the power metric, is currently the de facto industrial DC energy efficiency standard due to its simplicity. PUE is the ratio of total power used by the DC facility to the power used by the IT equipment.

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

Several case studies that have been conducted to analyse the PUE of DCs [16], [23]. The Data Center Infrastructure Efficiency (DCiE) [4], [5] is a metric initially recommended by The Green Grid, and it is synonymous with PUE, and it is the reciprocal of PUE.

$$DCiE = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}}$$

DC productivity metric is a crucial missing link in the energy efficiency issue [24]. The Green Grid [25] designed a family of Data Center Productivity (DCP) metrics for DC resource optimization. Its first derivative metric is Data Center Energy Productivity metric (DCeP), used to track the “*overall work product of a data center per unit of energy expended to produce this work*”. It is the ratio of useful work produced by the DC to the energy that is consumed to perform the work [26]. A new energy efficiency metrics for data centers introduced in [27] is based on the first law of thermodynamics thermal efficiency definition where thermal efficiency equals the ratio of useful work to total energy expended to operate and support the execution of useful work. However, the challenge is to define “useful work” in a meaningful, and measurable way [28] and in this paper, we demonstrate how this could be done (further discussed in Section III). Some productivity measurements related to useful work made by [29], [30] are the sum of weighted tasks carried out in a period and are represented in diverse ways: useful work per energy consumption; useful work per physical space; useful work to cost of the DC. In summary, it is useful for distinguishing the different operational states of IT equipment within a DC, and at the same time, identify IT infrastructure configurations that could improve energy productivity.

In this paper, we link the server performance metric to the utilization of servers. Studies show that for the period 2006–2012, the average server utilization remained at 12–18% [13]. Underutilized resources will not only have a significant energy draw but also impact on DC capacity. Several examples of metrics related to server utilization are: IT Equipment Utilization (ITEU) [31] which is linked to the effective use of IT equipment; Server Compute Efficiency (SCE) [12] to determine whether or not a server is being used to provide primary services and Data Center Compute Efficiency (DCcE) is calculated by averaging SCE values from all relevant servers in a period of time; Compute Power Efficiency (CPE) [32]

which is a measure of the computing efficiency of a DC and is defined as below:

$$CPE = \frac{(IT\ Equipment\ Utilization \cdot IT\ Equipment\ Power)}{(Total\ Facility\ Power)}$$

Lastly, green metrics are closely linked to sustainability and the environment. Typically, the green performance metric is defined as follows [33]:

$$Green\ Performance\ Metric = \frac{Useful\ Work}{Environmental\ Cost}$$

Green Energy Coefficient (GEC) metric involves the use of green or renewable energy [34]. It is the ratio of green energy consumed by a DC to its total energy consumption. Energy is dissipated by IT equipment in a data center. Energy Reuse Factor (ERF) [35], [36] is a metric for tracking the amount of waste energy reuse for sustainability purposes. The works [29], [30] have discussed a multi-dimensional approach for a new family of data center metrics which encompasses the following: performance-productivity, efficiency, sustainability and operations; risks. They have also highlighted that “downtime” (i.e. actual downtime, in terms of length, frequency, and recovery time) has an impact on productivity and ought to be measured as “useful work” that is not carried out due to failure or outage in the system. In the same vein of discussion, this paper looks into incomplete “useful work” where jobs are aborted and the measurement that we have created for this will be “waste energy” assessment.

III. METHODOLOGY

The work focuses on the evaluation of Productive Energy Consumption [9], [37], [38] by the ENEA HPC-DC, on the Cluster named CRESCO4 (hosted by ENEA-R.C. Portici), with the aim to explore a quantitative way to measure useful work (or “work done”) within the DC-Cluster and obtain qualitative interpretation based on the corresponding energy consumption. In this work, steps forward have been made in terms of the energy measurement and policies related to the power management with respect to the previous study conducted on the data available on CRESCO4 [32], [39]–[41]. In particular, the authors provide assessment of the useful work using in a practical way the Data Center energy Productivity (DCeP) metric and the waste energy estimation in DC, thanks to repartition of the waste energy. Computations necessary for the assessment are performed with the help of Python programming language, suitable for big datasets.

A. Dataset and Data Center Characteristics

Current analysis is based on available data from CRESCO4 that covers power and job schedule characteristics running on a cluster for 12 months (from February 2017 to February 2018). Hence, the analysis correlates accounting data from the Platform LSF (Load Sharing Facility) job scheduler and the corresponding energy consumption (from installed PDUs) obtained from data stored by Zabbix. LSF is a workload

management platform and job scheduler, for distributed HPC systems. This platform is concerned with deciding which process is to be run and is designed to keep CPUs as busy as possible. It stores a log file that contains all information on executed jobs and the usage of computing nodes (cores). Our experiment uses the LSF log file to understand the usage of the cores by the users and the Zabbix database that contains consumption values detected by the PDUs of the ENEA DC during job execution.

In detail, the experimental campaign evaluates the global energy consumption of the cluster CRESCO4 that consists of 38 Supermicro F617R3-FT chassis, each hosting 8 dual CPU nodes. Each CPU, specifically an Intel E5-2670, hosts in its turn 8 cores, for a total number of 4864 cores. These operate at a clock frequency of 2.6 GHz. Moreover, the system is provided with a RAM memory of 4 GB per core. Computing nodes access a DDN storage system, for a total storage amount of 1 Pbyte. Computing nodes are interconnected via an Infiniband 4xQDR QLogic/Intel12800-180 switch (432 ports, 40Gbps).

More specifically, two datasets have been available. The first dataset, Zabbix contains average level of power consumption, minimum and maximum registered power consumption for each hour between 12:00, 19th of February 2017 and 14:00, 19th of February 2018. The second set of data, LSF, covers details about number of cores assigned by the scheduler for every process, start and end time of the application activity, names of executable file and directory and the marker of whether the process has finished successfully (“done”) or with an error (“exit”). The second dataset covered IT jobs running within the interval from 21:12, 28th of December 2016 to 04:12, 25th of January 2018. From these datasets the time intersection was taken and covered 11 months from 12:00, 19th of February 2017, to 12:00, 25th of January 2018, divided by 19th day, 12:00, of each consecutive month except January 2018. The resulting datasets have covered over 8700 rows in Zabbix and 530668 rows in LSF datasets.

Regarding the scheduling of tasks in the DC-Cluster, the jobs are stored in the workload queues where the priority of each job and records them is processes according to the First Come First Served (FCFS) — a basic scheduling policy in which tasks are served in the order of their arrival in the system. This strategy reduces the waiting time of tasks.

The cluster have processed jobs from 18 different queues, 11 of which are composed of jobs to be performed with parallelization, three are considered with exclusively serial jobs and the remaining — with both parallel and serial jobs. Around 92% of all submitted jobs have been processed in serial mode, which leaves room for only 8% of jobs being calculated with parallelization techniques. The queues characteristics are reported in Table I.

The plethora of applications (approximately 40 types) that are running on the CRESCO4 cover several fields of research, such as materials science, efficient combustion, climate research, nuclear technology, plasma physics, biotechnology, aerospace, complex systems physics, renewable energies, en-

TABLE I
QUEUES CHARACTERISTICS

Queue name ^a	Number of submitted jobs	Queue duration (minutes)
SP: hpc_ha_h24	455764	1440
SP: cresco4_h6	25210	360
S: cresco4_h144	14738	8640
SP: small_h144	6160	8640
P: cresco4_256h24	3660	1440
P: cresco4minni_16h24	3267	1440
P: metinet	2860	1440
S: hpc_ha_h3	2266	180
S: small_10m	1626	10
SP: system	1576	6000000
P: cresco4_16h24	1565	1440
P: cresco4open_256h24	864	1440
P: combustion4	812	14400
P: fsn4_16h24	657	1440
P: cpu_plus	140	1440
P: clima	127	14400
P: chimere	68	20160
P: fsn3_24h24	4	1440

^aS: serial, P: parallel, SP: serial and parallel

vironmental issues, HPC technology. Moreover, many other kinds of applications are embedded in scripts that, through libraries, recall the functionality of consolidated software suites.

B. Analysis

As aforementioned, the goal of the work was to estimate effective energy consumption by the DC's cluster and identify levels of energy which IT jobs consumed. For this purpose, a following system of equations was proposed for each month:

$$\sum_{i=1}^K \int_{t_{i,j}^0}^{t_{i,j}^1} c_{i,j} \cdot x_j dt = E_j, \quad j = 1, \dots, N, \quad (1)$$

where x_j denotes the power required by any application every second from one core during the hour j , which is the principal unknown set of variables. Here i is an id of an active process; K stands for the number of processes registered to be active during the month under consideration; $c_{i,j}$ refers to the number of cores required to process application i during the current hour j ; variables $t_{i,j}^0, t_{i,j}^1$ correspond to the starting and ending

moment of the process i activity; E_j denotes energy consumed by the cluster during hour j in watt-hour; N is the number of hours in the extracted month. The left part of (1) represents a sum of integrals over the time of the processes activity. Expression under the integral sign represents power consumed by each process as a multiplication of unknown variable x_j and calculable coefficient $c_{i,j}$. This equation is then interpreted from integral to discrete format and is rewritten as follows:

$$\sum_{i=1}^K c_{i,j} \cdot \frac{t_{i,j}}{3600} \cdot x_j = E_j, \quad j = 1, \dots, N, \quad (2)$$

where $t_{i,j}$ describes the duration of process i activity in seconds, explaining the need to divide it by the number of seconds in one hour. Thus, the sequence of linear equations each having one unknown is formed and can be resolved simply dividing the right part by the sum from the left part.

Obtained solution is further filtered to exclude the outliers, which possibly appeared as a result of missing data. The outliers have been identified by the following formula:

$$|x_j - \bar{x}| > 2 \cdot (Q_3(x) - Q_1(x)), \quad (3)$$

where \bar{x} stands for the mean of the vector $x = (x_j, j = 1, \dots, N)$. $Q_1(x), Q_3(x)$ denote the first and the third quartiles of the vector x correspondingly, under assumption of x being normally distributed. Since quartiles are forth quantiles, $Q_1(x)$ and $Q_3(x)$ values separate a set of x values into four subsets of equal size. Thus, x_j is considered to be an outlier, if its Euclidian distance from the mean value is greater than twice the distance between two values of the two densest subsets, on which quantiles divided the values of vector x .

The sum on the left-hand side of (2) has been checked for the equivalence to zero. In such cases when the sum $\sum_{i=1}^K (c_{i,j} \cdot t_{i,j})$ equals zero, the hour j has been assumed to have insufficient data and is not considered for final inferences. The decision of excluding such hours from the resulting dataset is based on the trial of idle power estimation. Average energy consumption during the hours with active processes running is 47 kWh, while it equals 42 kWh when no processes are registered to be active, i.e. when the aforementioned sum equals zero. Moreover, the range of energy consumed, when no process has been active, lays within the range of energy consumption when cores have been reported to work on the jobs. Namely, if all hours with no processes registered to be active are united into one dataset, the range of energy consumption is [27.3; 58.8] kWh, whereas in the dataset with non-zero sums of coefficients for each hour the energy consumption lay within the interval of [14.4; 65.5] kWh. This inclusion does not allow making estimation on idle energy consumption, because it lies within the energy consumption range reported for the hours with active processes running.

It is also worth mentioning that a more granular data analysis could take place with the following system:

$$\sum_{i=1}^K c_{i,j} \cdot \frac{t_{i,j}}{3600} \cdot x_{i,j} = E_j, \quad j = 1, \dots, N. \quad (4)$$

It differs from the previous system (2), since the unknown variables $x_{i,j}$ here additionally correspond to each individual process i power consumption per core each second. The matrix of SLAE in (4) has the dimension $N \times M$, where M reflects the number of processes active during hour j , compared to the matrix of dimension $N \times 1$ in the previous system (2). High granularity system has been built to get more precise values, nevertheless, some characteristics of such system have prevented from getting expected results.

Exemplar situation could be shown based on the system for the month of 19th February – 19th February 2017. The matrix of the system is sparse and has only 3% of non-zero elements. The SVD analysis of the system has shown that conditional number of the matrix has the order of 10^{16} . These properties together with non-negativity constraint for $x_{i,j}$ have resulted in poor accuracy of the solution obtained with the help of algorithms for ill-posed problems (Least Squares, Least Squares with regularization in \mathbb{L}_2 , Least Squares with regularization in \mathbb{L}_2 and \mathbb{L}_1). For this reason, a decision of decreasing granularity of unknown variables has been made in order to obtain results, which represent the average multiple x_j of power consumption for all jobs registered for the hour j . Once the vector x_j is obtained, multiplying elements of this vector and a corresponding weight can derive energy consumption of individual processes ($c_{i,j} \cdot t_{i,j}/3600$).

The mathematically formulated analysis will allow investigating on DC energy metrics since its result is expressed in terms of energy used by every application each month (with a granularity related to each hour). It is then crucial to assess what portions of energy have been spent on “useful work” and have accumulated in “waste energy”. alternatively from every application to draw a clear picture of energy profiles and energy waste distribution by identified categories and enable suitable improvement actions for the DC. This estimation will be done with productivity related metrics introduced in the next subsection.

C. Productivity Related Metrics

The clusters energy consumption is evaluated with energy efficiency metric as well as categorization of waste energy sources. DC energy productivity (DCeP) metric [42], [43] involved in analysis is expressed as follows:

$$DCeP = \frac{Useful\ Work\ Produced}{Total\ DC\ Energy\ Consumed\ over\ Time}. \quad (5)$$

The available data allow defining *Useful Work Produced* as energy spent on jobs which processing ended correctly. *Total DC Energy Consumed over Time* is represented here by the energy used for all jobs, both properly and incorrectly ended. Although generally accepted practice is to consider energy, which goes for both cooling and IT systems, under the notion of *Total DC Energy Consumed over Time*, limitations of the data retrieved from the cluster do not provide sufficient information to study.

D. Waste Energy Analysis

One way to reduce overall energy consumption in DC is to understand the Energy Waste causes and then to improve the power management thanks to the workloads analysis. For this reason, the work also focuses on categorizing submitted jobs to distinguish between the jobs, which resulted in the effective work done, and the jobs (or the “not jobs”) that represent only inefficient work with their inefficient energy use. The latter jobs can be subdivided into three categories to assess their contribution to wasted energy:

- a) Jobs with maximum running time of 30 seconds;
- b) Jobs which exceeded the queue time;
- c) Jobs which left the queue with an error for any other reason.

The a) category comprises jobs namely “not jobs” with such short running time, which occurred to represent the work of the scheduler, while the scheduled application itself has not been started. The value of the threshold at 30 seconds is an empirical choice due to the knowledge of the preworking time of the LSF application and then its dataset. The jobs running during less than 30 seconds represent the preprocessing phase and they have not produced any useful work in terms of results for the end-user, who submitted them to the cluster. For this reason these jobs are considered to cause waste energy. Given that most of the jobs running time has varied from two seconds to 221 hours, the average being two hours, the processes from category a) have consumed small amounts of energy. However, the presence of such processes affects the cluster work.

The group b) consolidates all jobs the running time of which exceeded the queue maximum time. The existing policy of the DC usage states that if a job is submitted to processing units, and allocated into a specific queue by LSF, the queue allows this job to run for a certain time. In case of exceeding the maximum time assigned by the queue, the job is removed from the queue, being reported as erroneous process, sometime after the queue time limit is exhausted. However, while the job is being processed within the queue time, it produces results and cannot be regarded as a reason for waste energy. For example, taking into account a job with exit status that was running on the queue lasting 600 seconds. If it started at 1494316196 (unixtime) and ended at 1494317094 (unixtime) on this kind of queue, the total job duration (stop-start) is 898 sec and exceeded the max queue time. According to the empirical data and the cluster workload-policies (related to some applications), clarification and explanation have been provided. Hence, in our analysis, we calculate as useful work the work associated with the part of the job that was running for total queue duration (600 sec); meanwhile, we consider as waste energy the part of energy spent to the job running in the rest of the time $(898 - 600) = 298$ sec. The energy is wasted only when the job runs after the queue time limit, which is the focus of the b) category.

The third category c) is composed of jobs with any other malfunction causing jobs’ interruption both by end-users or by the system.

The three defined categories are further used to measure Energy Wasted Ratio (EWR) [42], [44]:

$$EWR = \frac{\text{Energy Wasted for not Useful Work}}{\text{Total DC Energy Consumed over Time}}. \quad (6)$$

This metric assesses ratio of energy spent on the work which has not provided any useful result, or on the “not jobs” from the energy waste categories, which will be covered in the next section.

IV. RESULTS & DISCUSSION

An overview of all computational process present in the DC Cluster is shown in by the previous section with aims to provide a further step in the assessment of the DC energy efficiency scenario. This section is devoted to the discussion to the results achieved by the experimental campaign based on the big data set obtained during one year of monitoring. The power management strategies and policies devoted to DC operators and costumers — based on the statistical data analysis — can be identified and suggested. Moreover, the quality of the power management strategies [45] is suggested based on both power and performance measurement data collected during one year of the system operation.

The Fig. 1 provides an indication related to the efficiency of the workloads management. Indeed, the figure illustrates amounts of energy consumed by applications working on various tasks during the overall period in question. The largest share of cluster’s energy consumption was dedicated for simulations of particles detection with help of Monte Carlo algorithms as well as for air quality research (FARM+RAMS, MINNI) and climate modelling (REGCM4). The smallest portion of energy was used for a type of Monte Carlo analysis, genetic analysis and mathematical algorithm (HETERO3D). All the applications which required less than 1% of annual energy individually were combined into one group. The ratio of energy used in total by this group comprises 16% of the yearly energy, which again reveals an energy pattern when a considerable number of low-energy consumption applications are processed on the cluster.

A. Assessment of DCeP

As mentioned in the introduction, until now in the DC metrics scenario limited attention has been given to the development of metrics aimed at clearly defining the “useful work”, that is, metrics aimed to gauge the real “computing” carried out by a DC. Generally, the useful work of a DC is represented by the computing activity of IT (computing, storing and transferring data — IT Services) and appropriate productivity metrics can measure it [9], [43]. Productivity metrics differ from other metrics in their approach to how they assess useful work, however, even if several attempts have been made to define the productivity metrics for DCs none of the metrics have provides the practical way to calculate the effective work done. In the productivity metrics scenario [46], the most significant is the DCeP metric, and our data analysis has been devoted to calculating and using in our experimental campaign.

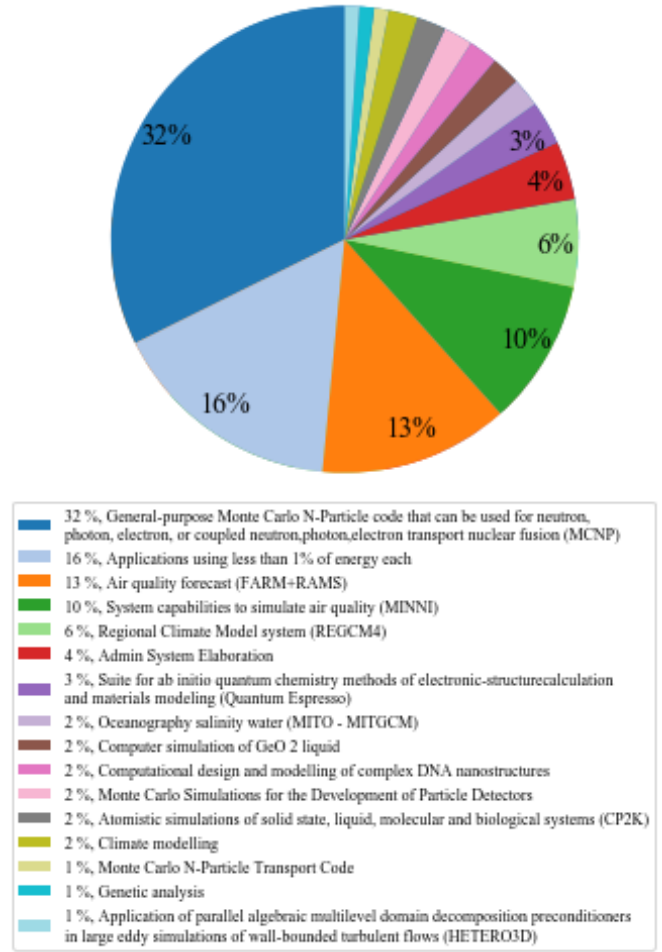


Fig. 1. Overall energy use by applications.

To calculate the formula we started from the job analysis because the knowledge of their scheduling and management is an essential part of the DC configuration and it influences the energy consumption and the performance a lot.

The estimation of every IT jobs energy consumption has allowed evaluating the load on the DC cluster, as well as efficiency of energy use. Monthly energy consumption by each application with separation on the jobs, which finished with error and those finished correctly, are obtained for every month of the final datasets for the 11 months duration.

Based on the estimation of each IT job power consumption per core during each second from (2) and given the information about fulfillment status of jobs, it is possible to estimate DCeP metric. The Fig. 2 represents energy consumption for processes, which ended with an error and those, which were successfully completed. The largest portion of energy use was observed during the month from 19th of March to 19th of April reaching the point of 35.6 MWh, whereas the smallest portion energy consumption was reported during the months from 19th of July to 19th of September. DCeP varies from 0.61 in the last reported month to 0.84 in June–July period.

In the case when jobs were not categorized on three energy

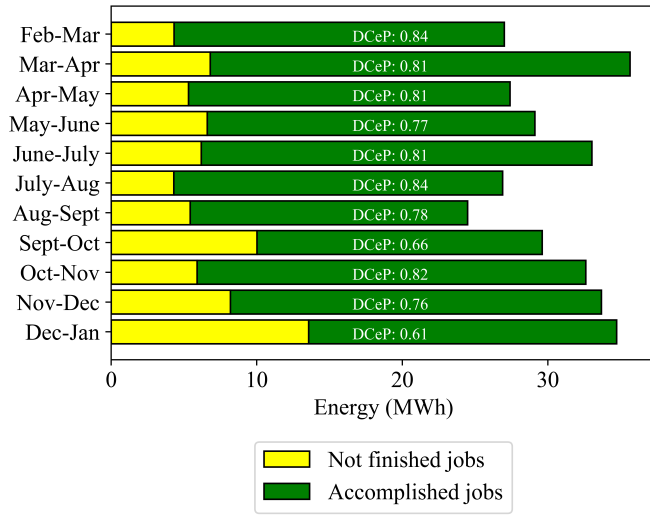


Fig. 2. Monthly analysis of energy consumed by correctly finished jobs (“useful work”) and jobs which exited a queue with error status (“energy waste”) and DCEP. Energy waste categories are considered.

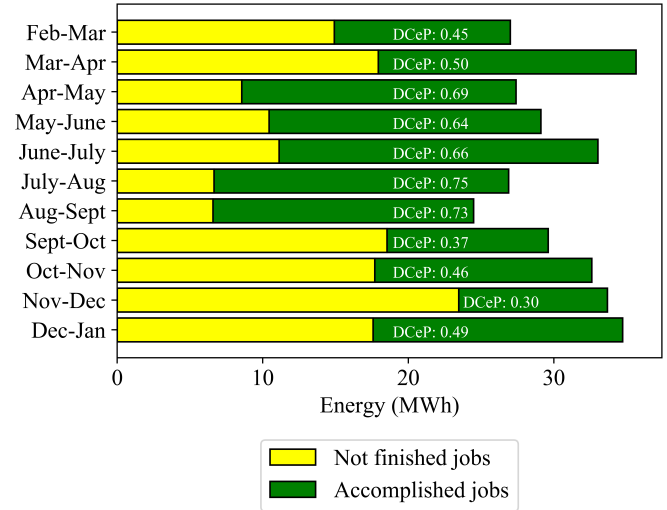


Fig. 3. Monthly analysis of energy consumed by correctly finished jobs (“useful work”) and jobs which exited a queue with error status (“energy waste”) and DCEP. Jobs are not categorized by causes of energy waste, data on jobs status is taken directly from LSF.

waste groups, but were taken directly from the LSF data, DCEP was reported to stay at lower level than after preprocessing the LSF dataset and extracting categories. The DCEP differences can be observed in the Fig.3 where no categorization has been done versus the Fig. 2 depicting values when categorization has been taken into account. These results are explained by the fact that in the raw LSF dataset all the jobs exceeding the queue time were marked with “exit” status and counted as energy waste, whereas as described previously, the energy used within the queue time had been spent on useful work and only the remaining part of processing period caused energy waste. In addition, some jobs with duration time within 30 seconds were marked as useful work which, according to our assumptions, is not true. Henceforward, the categorized dataset is used, i.e. the one corresponding to the Fig. 2.

Energy consumption of the processes is found to be unevenly distributed. The majority of the processes have consumed less than 100 kWh per month. More granular analysis shows that from 62% to 93% of overall number of IT jobs running in the cluster consume less than 10 kWh per month (Fig. 4).

B. Assessment of Waste Energy

Since the evaluation of the useful work is linked to the waste energy in computational terms, an assessment of the “waste” has been made. Hence, job distribution into waste energy categories is analyzed both from the point of their energy consumption and their share in submitted jobs number. In this sub-paragraph, we investigate on the “not useful work” to calculate how much energy using for computing activities is wasted and is not used for producing the useful work.

IT jobs, which had been running on the cluster, were distributed into energy waste categories as explained in section III. Statistical characteristics are taken from the monthly

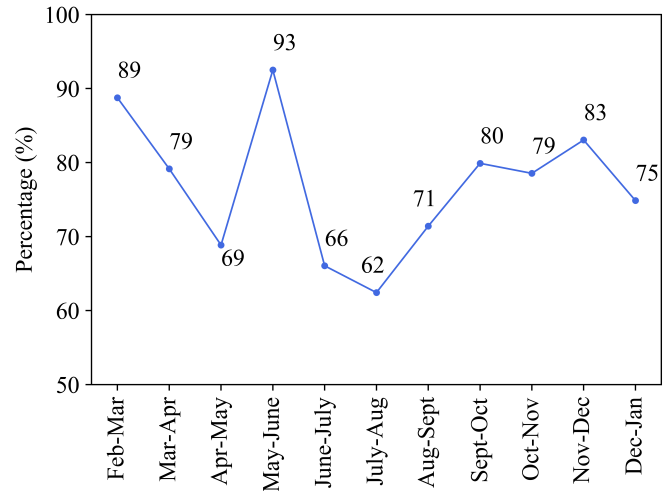


Fig. 4. Ratio of executed jobs which consumed less than 10 kWh per month and ended either with or without an error in correspondence to overall number of jobs.

TABLE II
ENERGY WASTE CATEGORIES

Statistical Characteristics	Energy Waste Ratio by Job Categories with Relation to Overall Energy Use (%)		
	a) Running time ≤ 30 sec	b) Running time $>$ queue time	c) Other reasons
Min	$7 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	16
Max	$6 \cdot 10^{-2}$	0.3	39
Mean	$3 \cdot 10^{-2}$	0.2	23
Standard Deviation	10^{-2}	$9 \cdot 10^{-2}$	7

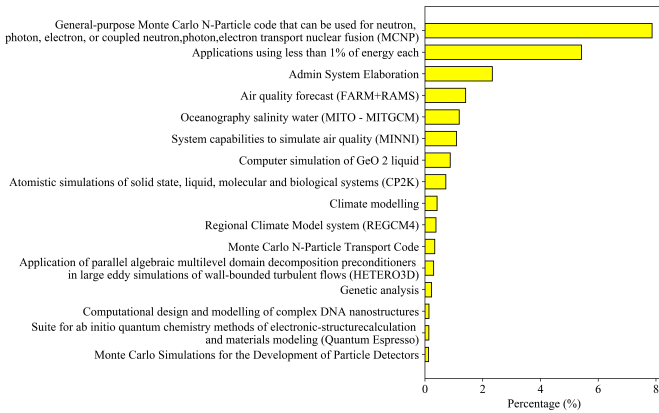


Fig. 5. Energy Waste Ratio (EWR) by applications.

samples of data and are shown in Table II, namely minimum, maximum, mean value and standard deviation of the ratios of energy used by jobs from each category related to the general energy use. As might be observed from the Table II, processes with short running time consumed the least share of energy, 0.03% of overall energy use, whereas jobs exceeding the queue time used around 0.2% of energy. Both categories have a small deviation from the mean value, which signifies moderate fluctuation of their energy use. By contrast, processes malfunctioning for different other reasons used a range of 16 to 39% of energy. The deviation of the latter category is the highest, which highlights the necessity of closer investigation of erroneous jobs processing to decrease the number of their submissions and increase energy productivity of the cluster. Values of EWR for all the three categories combined together are associated with applications and represented in the Fig.5 in percents. When compared with the Fig.1, the pattern is similar: Monte Carlo analysis has consumed the most of the energy over 11 month causing the largest energy waste of all applications, as well as applications having required small amounts of energy create correspondingly low energy loss.

Monthly analysis dealing with the categories of jobs having caused energy waste is characterized by Fig. 6 representing numbers of submitted jobs which resulted in a) and b) categories concerning the number of all active jobs. A considerable amount of jobs which were only processed by the scheduler and had a maximum running time of 30 seconds (category a)) is reported throughout the whole period of investigation, composing from 14 to 56% of all submitted jobs. On contrary, jobs, which exceeded the queue time limit, formed less than 1% during the majority of reported period. Seasonal steep rise of jobs submission from both categories happened in the month of November-December, which might be associated with users working from home during Christmas vacations and failing to conform to the DC policies.

To summarize, results obtained through the analysis reveal the energy consumption patterns within the cluster. Firstly, the least energy was consumed during the summer months of annual vacations, whereas the most significant amount of

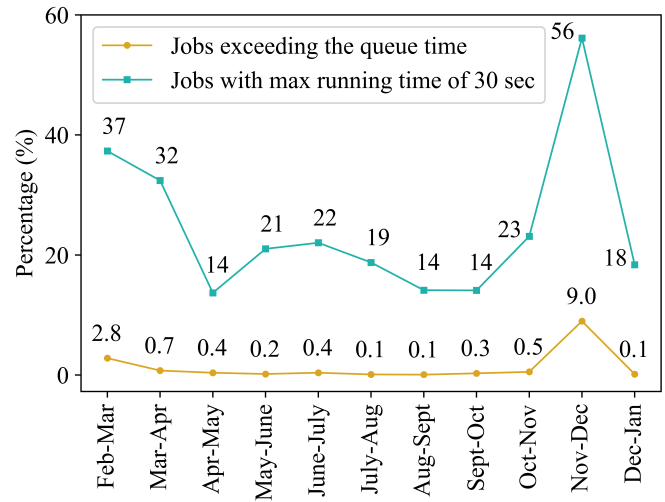


Fig. 6. Ratio of jobs causing energy waste because of processing time conflicts.

wasted energy was observable in December-January when end users might have worked remotely due to the Christmas holidays. Secondly, a high percentage of jobs consume less than 10 kWh per month, which results in the energy being spent on small jobs rather than resource-consuming processes. Also, the cluster wastes the most energy for erroneously ended jobs for unknown reasons, which should be further examined. Regarding the waste energy categories, some jobs that were only preprocessed by the scheduler and did not provide any results, is considerably higher than the number of jobs removed from the queue because of the time limit conflicts.

C. Definition of General Policies & Strategies

The power management strategies presented in this sub-paragraph, could significantly reduce the energy consumption in the CRESCO4 cluster but also can be generalized for a typical HPC DC workload with the same characteristics.

At the beginning to improve the energy consumption and then the power management we decide to start and distinguish three fundamental component related to the resource management and scheduling policies from the DC operator side: 1) scheduling, 2) resource allocation and 3) resource management.

The scheduling is responsible for defining the order of execution for the active tasks, and as aforementioned the strategy used in our case study is based on the FCFS algorithm. However, taking into account, the results obtained and the potential of LSF platform for job-scheduler other strategies can be adapted to define the order in which tasks are placed in the queues. In literature, some widely used algorithms and related policies can be applied to the scheduling of tasks in DCs [8]. For example, the Last Come First Served (LCFS) — the Largest Job First (LJF) with the aim to optimize the utilization of the system — the Smallest Job First (SJF) to increases the throughput of the system. The policies referred

to scheduling algorithms can be extended with one of the backfilling approaches (such as, Aggressive Backfill, Relaxed Backfill, Conservative Backfill) which exchange the positions of the jobs in the queue based on the availability of the resources and the priorities of the tasks.

The current resource allocation used by the costumers is based on the “free queue using” and then on the “on time” strategy: “the first user entry in the system using the first available requested queue”. However, to define the manner in which tasks are assigned to resources — since the tasks are submitted by different users over time — deciding of where to execute each arriving task is usually made in an online manner without knowledge of the future task arrivals is needed. The performance-energy combination is the driver of the allocation strategy which aim is to minimize the average task response time, the overall energy consumption. The tasks can also be assigned to resources to consolidate the workload in a predefined allocation manner. The primary strategy regarding the Energy Usage Optimization thanks to each job is assigned to a node that minimizes the energy consumed by the task.

The latter point regarding the resource management policies that specify a set of operations performed on the resources during the scheduling process. They usually require support from the underlying hardware layer and their effectiveness is closely related to the managed IT equipment. The most popular policies are the Switching nodes ON/OFF and the Dynamic Voltage and Frequency Scaling (DVFS).

According to the CRESCO4 results, the previous three components (scheduling, resource allocation and resource management) can be combined in ad-hoc policies for the power management. There are several ways to follow to get “workload driven” power management. First of all, the authors pursue the power cost optimization. From LSF information it can be seen that the majority of job requests arrived in the morning during working hours. Generally, it is the time when energy cost is high due to power cost changes throughout a day. To high demand corresponding to a high price vice-versa, low demand suggests a low price. The power cost optimization strategy is used to minimize power cost by shifting the workload to low-cost periods (when it is possible or when company policies allow it). In the case where system utilization is close to 100%, an alternative strategy is to schedule only high priority work during the highest power cost period and prevent workload which can wait (i.e. low priority) from consuming power when it is most expensive.

From the results obtained it is easy to understand that the system is not always 100% charged and a lot of time passes with the system loaded only in half or even less. It is also deduced that the objective of concentrating the load is not pursued because waiting time for the jobs is minimized in the first place leaving the energy consumption minimization as lower priority. The power consumption of a node depends on its operational mode. When the node is off, and in sleep/standby mode, it consumes very little power. If a node is idle, it consumes 40–50% of power compared to a fully loaded node. For the power efficiency optimization

strategy to work, we first have to understand the performance of a particular class of server per kilowatt. Such an efficiency metric can be application-dependent and therefore should be considered carefully. Using tools such as Platform LSF one can easily manage the distribution of loads on the various nodes.

During the analysis, we detected some applications requiring heavy I/O or, applications that wait for MPI messages, can be classified as low load or cooling load because the CPU tends to go idle and consume less power during these waiting periods. The strategies mentioned above can be replicated in the other DCs Cluster with the same; hence, it is possible to redistribute the jobs to optimize the energy consumption and also taking into account the thermal awareness [47].

D. Sustainability

As observed in Section IV.C, the DC cluster experiences peaks of energy consumption in the morning hours, which, apart from increasing energy cost affect the amount of carbon emitted as a consequence of electricity generation. Two ways of reducing the harmful impact of energy consumption peaks might be brought into action: 1) executing only “high priority work during the highest power cost period” 2) equip DC with their sources of renewable energy like solar panels and windmills to further adapt the execution of works regarding the local renewable energy production [48].

Further considering the sustainable issue, the ISO standard 30134 being developed for sustainable IT and DCs includes a renewable energy factor as a key performance indicator for DCs [49]. However, others ISO regarding IT and energy efficiency and sustainable topics in DC are currently in development. Even if the ongoing innovation actions (Reduce/Reuse/Recycling) help to drive the DC towards becoming genuinely sustainable, poorly utilized of the defined metrics related to the IT and then the useful work, will impede business innovation and prevent meeting environmental sustainability goals. Also, for this reason, in this work, the authors have investigated the practical application to the case study of the EWR metric (6). In detail, the analysis allows precisely to pinpoint the non-computing and non-useful work within the energy waste. Therefore, based on the knowledge of the energy waste, it is possible to improve the power management regardless of the job consumption and to pursue the sustainable issue concerning the IT reduction action for the job.

V. CONCLUSIONS & FUTURE WORK

Analysis and measurement of the DC energy data are the baselines to achieve the energy efficiency goal; the metrics are essential tools for the monitoring of energy use and environmental conditions at a DC scale. These metrics make it possible to quantify the consumption, evaluate the performance of a DC and compare different technologies and strategies. De facto, a highly efficient DC is valuable regarding owners and operators and appeal to customers. Moreover, the need to invest in energy efficiency is becoming a priority in the Smart

Cities context. In this paper, we have presented the methodology and assessment of a specific class of power/energy metrics for DC, the productivity metrics with the aim of making the computation of IT useful work (related to the computing work in DC) more profound and real knowledge. Even if the results on productivity metrics for predicting energy IT consumption are ongoing to develop, however, their adoption and application to the real case is a challenge. For this reason, this work highlights the importance of the enactment of productivity metrics related to useful work and investigate on the energy consumption when the applications were running. The assessment of the productivity metrics and the waste energy evaluation carry out from a big data set (that cover 12 months) of ENEA DC, Cluster. We have presented workload profiles, application-workloads, power and energy model approach to identify the useful work and non-work in the cluster. Besides, productivity and waste energy metrics at the application level of analysis were proposed. Various resource management and scheduling policies, including performance, energy and consolidation policies were presented. It can be easily seen that introducing waste energy assessment to suggest the power management strategies which will improve the DC' efficiency significantly. The real data set allows validating the metrics and the assessment to improve the QoS of DC. Additional insight into the sustainability issue is considered concerning the reduction to the non-work coming from to the estimates factor of energy that is wasted in the cluster.

In future works, the prediction models based on energy consumption of applications and thermal data set at room and node level will be conducted. We will also use the power management strategies to reduce the energy consumption will apply. To validate the policy and the prediction models the real case will be conducted on the new ENEA Cluster, CRESCO6.

ACKNOWLEDGMENT

The research work has been supported and funded by the PERCCOM Erasmus Mundus Program of the European Union [50]. Moreover, the authors would like to express their gratitude to the research HPC group at the ENEA-R.C. Portici for the useful advice on modelling and control of ENEA-Data Center.

REFERENCES

- [1] Shehabi, A., et. al. (2016). United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-1005775, url: http://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf.
- [2] European Commission. (2016). Code of Conduct for Energy Efficiency in Data Centres, url: <https://ec.europa.eu/jrc/en/energy-efficiency/code-conduct/datacentres>.
- [3] European Commission. (2018). 2018 Best Practice Guidelines for the EU Code of Conduct on Data Centre Energy Efficiency, Version 9.1.0, url: <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC110666/kjna29103enn.pdf>.
- [4] BCS. (nd.a). Data centre energy efficiency metrics: Existing and proposed metrics to provide effective understanding and reporting of data centre energy, url: <https://bcs.org/upload/pdf/data-centre-energy.pdf>.
- [5] The Green Grid. (2007). The Green Grid Data Center Power Efficiency Metrics: PUE And DCiE, url: http://www.premiersolutionsco.com/wp-content/uploads/TGG_Data_Center_Power_Efficiency_Metrics_PUE_and_DCiE.pdf.
- [6] Zoie, R. C., et. al. (2017). An analysis of the Power Usage Effectiveness metric in Data Centers, Proceedings of 5th International Symposium on Electrical and Electronics Engineering (ISEEE), pp. 1 - 6, doi: 10.1109/ISEEE.2017.8170650.
- [7] Van De Voort, T., et. al. (2017). Analysis of performance metrics for data center efficiency should the Power Utilization Effectiveness PUE still be used as the main indicator Part 1? REHVA Journal February 2017.
- [8] Van De Voort, T., et. al. (2017). Analysis of performance metrics for data center efficiency should the Power Utilization Effectiveness PUE still be used as the main indicator Part 2? REHVA Journal, 2017(2), pp.:37-43.
- [9] M. Chinnici, A. Capozzoli, G. Serale, Measuring energy efficiency in data centers, in book Pervasive Computing Next Generation Platforms for Intelligent Data Collection, Chapter 10, pp. 299-351, 2016.
- [10] Brill, K. G. [2007]. Data center energy efficiency and productivity, <http://large.stanford.edu/courses/2017/ph240/you2/docs/brill.pdf>.
- [11] The Green Grid. [2008]. A Framework for Data Center Energy Productivity, url: <https://www.greenbiz.com/sites/default/files/document/GreenGrid-Framework-Data-Center-Energy-Productivity.pdf>.
- [12] The Green Grid. (2010b). The Green Grid Data Center Compute Efficiency Metric: DCcE, White Paper #34, url: <http://docshare03.docshare.tips/files/14778/147781031.pdf>.
- [13] NRDC and Anthesis. (2014). Data Center Efficiency Assessment: scaling up energy efficiency across the Data Center Industry, evaluating Key Drivers and Barriers, Issue Paper August 2014 IP:14-08-a, url: <https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf>.
- [14] Baccour, E., et. al. (2017). PTNet: An efficient and green data center network, J. Parallel Distributed Computing, Vol. 107, pp. 318.
- [15] Intel. (2010). Increasing Data Center Efficiency with Server Power Measurements, url: <https://www.intel.com/content/dam/doc/white-paper/intel-it-data-center-efficiency-server-power-paper.pdf>.
- [16] Pattinson, C., Cash, G., Kor, A. L., and Cross, R. (2015). Critical Issues for Data Center Energy Efficiency. In Dastbaz, M., Akhgar, B., and Pattinson, C. (eds.). Green Information Technology: A Sustainable Approach, Morgan Kaufmann.
- [17] Lajevardi, B., Haapala, K. R., and Junker, J. F. (2015). Real-time monitoring and evaluation of energy efficiency and thermal management of data centers, Journal of Manufacturing Systems, Vol. 37, pp. 511-516.
- [18] Sisk, D. R., et. al. (2009). Real-Time Data Center Energy Efficiency at Pacific Northwest National Laboratory, ASHRAE Transactions; Atlanta Vol. 115, (2009), pp. 242-253.
- [19] Lin, W., et. al (2018). Experimental and quantitative analysis of server power model for cloud data centers, Future Generation Systems, Vol. 86, pp. 940-950.
- [20] Daim, T., et. al. (2009). Data center metrics: An energy efficiency model for information technology managers, Management of Environmental Quality: An International Journal, Vol. 20 Issue: 6, pp.712-731, <https://doi.org/10.1108/1477830910990870>.
- [21] Yuventi, J., and Mehdizadeh, R. (2013). A critical analysis of Power Usage Effectiveness and its use in communicating data center energy consumption, Energy and Buildings, Vol. 64, pp.90-94.
- [22] Whitehead, B., et.al. (2014). Assessing the environmental impact of data centres part I: Background, energy use and metrics, Building and Environment, Vol. 82, pp.151-159.
- [23] Brady, G. A., Kapur, N., Summers, J. L., and Thompson, H. M. (2013). A case study and critical assessment in calculating power usage effectiveness for a data center, Energy Conversion and Management, Vol. 76, pp.155-161.
- [24] Schaeppi, B., et. al. (2012). Metrics for energy efficiency assessment in data centers and server rooms, Proceedings of Electronics Goes Green 2012+ Conference, 9-12 Sept. 2012, Berlin: Germany.
- [25] The Green Grid. (2008). A Framework for Data Center Energy Productivity, White Paper #13, url: <https://www.greenbiz.com/sites/default/files/document/GreenGrid-Framework-Data-Center-Energy-Productivity.pdf>.
- [26] Sego, L. H. (2012). Implementing the Data Center Energy Productivity Metric, ACM Journal on Emerging Technologies in Computing Systems, Vol. 8, No. 4, Article 30, Pub. date: October 2012.
- [27] Beitelmal, A. H., and Fabris, D. (2014). Servers and data centers energy performance metrics, Energy and Buildings, Vol. 80, pp. 562-569.

- [28] Horner, N., and Azevedo, I. (2016). Power usage effectiveness in data centers: Overloaded and Underachieving, *The Electricity Journal*, Vol. 29, pp. 6169.
- [29] Levy, M., and Raviv, D. (2017). A Novel Framework for Data Center Metrics using a Multidimensional Approach, *Proceedings of 15th LACCEI International Multi-Conference for Engineering, Education, and Technology: Global Partnerships for Development and Engineering Education*, 19-21 July 2017, Boca Raton FL, United States, DOI: <http://dx.doi.org/10.18687/LACCEI2017.1.1.387>
- [30] Levy, M., and Raviv, D. (2018). An Overview of Data Center Metrics and a Novel Approach for a New Family of Metrics, *Advances in Science, Technology and Engineering Systems Journal* Vol. 3, No. 2, pp.238-251 (2018).
- [31] Green IT Promotion Council. (2012). New Data Center Energy Efficiency Evaluation Index DPPE (Datacenter Performance per Energy) Measurement Guidelines, url: https://home.jeita.or.jp/greenit-pe/topics/release/pdf/dppe_e_DPPE_Measurement_Guidelines.pdf.
- [32] M. Chinnici, A. Quintiliani, "An example of methodology to assess energy efficiency in Data Centers", in *IEEE Third Int. Conf. Cloud Karlsruhe*, pp. 459-463, 2013.
- [33] Wang, L., and Khan, S. U. (2013). Review of performance metrics for green data centers: a taxonomy study, *Journal Supercomputing*, Vol. 63, pp. 639–656, DOI 10.1007/s11227-011-0704-3.
- [34] Shiino, T. (2012). Standardizing data center energy efficiency metrics in preparation for global competition, *NRI papers*, No 171, March 1, 2012, url: <https://www.nri.com/global/opinion/papers/2012/pdf/np2012171.pdf>.
- [35] The Green Grid. (2010a). ERE: A Metric for Measuring the Benefit of Reuse Energy from A Data Center, White Paper #29, url: https://eehpcwg.llnl.gov/documents/infra/06_energyreuseefficiencymetric.pdf
- [36] BCS. (nd.b). Future of data centre efficiency metrics: A review of progress so far and a suggested direction for further development, url: https://www.bcs.org/upload/pdf/future_of_data_centre_efficiency_metrics_v2.0.0_release.pdf.
- [37] A. Capozzoli, M. Chinnici, M. Perino, G. Serale, Review on Performance Metrics for Energy Efficiency in Data Center: The Role of Thermal Management, in *Springer E2DC 2014*, Cambridge, Vol. 8945, 135-151, 2015.
- [38] A. Capozzoli, G. Serale, L. Liuzzo, M. Chinnici, Thermal Metrics for Data Centers: A Critical Review, in *ENERGY PROCEDIA Elsevier*, Vol. 62, pp. 391-400, 2014.
- [39] M. Chinnici, D. De Chiara, A. Quintiliani, Data Center, a Cyber-Physical System: Improving Energy Efficiency Through the Power Management, in *Proceedings of 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 2017 IEEE 15th International Conference on Pervasive Intelligence and Computing, 2017 IEEE 3rd International Conference on Big Data Intelligence and Computing and 2017 IEEE Cyber Science and Technology Congress, DASC-PICom-DataCom-CyberSciTec 2017*, Vol. 2018-January, p. 269-272, 29 March 2018.
- [40] A. Quintiliani, M. Chinnici, D. De Chiara, Understanding 'workload-related' metrics for energy efficiency in Data Center, in *20th International Conference on System Theory, Control and Computing, ICSTCC*, pp. 830-837, 2016.
- [41] M. Chinnici, D. De Chiara, A. Quintiliani, An HPC-Data Center Case Study on the Power Consumption of Workload , in book *Applied Physics, System Science and Computers II*, Springer, Vol. 489, Pages 183-192, 2019.
- [42] L. Cupertino, et al., Energy-efficiency, thermal aware modeling and simulation of data centres: the CoolEmAll approach and evaluation results. *Ad Hoc Netw.*, Vol. 25, p. 535-553, 2015.
- [43] A. Quintiliani, M. Chinnici, et al., Final DC4Cities standardization framework and results description of the European Cluster, ENEA, Rome, Italy, WP7 Deliverable 7.3 , Project # 609304, 16 March, 2016. url: <http://www.dc4cities.eu/en/wp-content/uploads/2016/05/D7.3-Final-DC4Cities-standardization-framework-and-results-description-of-the-European-Cluster.pdf>
- [44] I. Munteanu, V. Debusschere, S. Bergeon, and S. Bacha, Efficiency metrics for qualification of datacenters in terms of useful workload, in *IEEE Grenoble Conference PowerTech, POWERTECH 2013*, 2013.
- [45] B.F. Postema, B.R. Haverkort, Evaluation of advanced data centre power management strategies, *Electronic Notes in Theoretical Computer Science*, Vol. 337, p. 173-191, 2018.
- [46] V. D. Reddy, et al., Metrics for Sustainable Data Centers, *IEEE Transaction on sustainable computing*, Vol. 2, N.3, 2017.
- [47] A. Capozzoli, G. Primiceri, "Cooling systems in data centers: State of art and emerging technologies", *Energy Procedia*, Vol. 83, p. 484-493, 2015.
- [48] N. Beldiceanu et al., "Towards energy-proportional clouds partially powered by renewable energy", *Computing*, vol. 99, no. 1, pp. 322, 2017.
- [49] ISO.org. (2018). [online] Available at: <https://www.iso.org/search.html?q=data%20centres> [Accessed 20 Jul. 2018]
- [50] A. Klimova, E. Rondeau, K. Andersson, J. Porras, A. Rybin, A. Zaslavsky, An international Master's program in green ICT as a contribution to sustainable development, *Journal of Cleaner Production*, Vol. 135, p. 223-239, 2016.