A Novel Framework for Data Center Metrics using a Multidimensional Approach

Moises Levy, MSc, Daniel Raviv, PhD
Florida Atlantic University, USA, mlevy2015@fau.edu, ravivd@fau.edu

Abstract– Given the mission critical nature of data centers, metrics must provide a holistic understanding of the data center behavior. After thoroughly reviewing existing data center metrics, this paper proposes a novel framework for more accurate metrics for data centers using a multidimensional approach. Performance is examined across four different sub-dimensions: productivity, efficiency, sustainability, and operations. Risk specifically associated with each of those sub-dimensions is contemplated. External risks are also considered independently of performance, namely location risk, global risks and site infrastructure risk.

Keywords—Data center, metrics, performance, efficiency, risk.

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2017.1.1.387
ISBN: 978-0-9993443-0-9
ISSN: 2414-6390

A Novel Framework for Data Center Metrics using a Multidimensional Approach

Moises Levy, MSc, Daniel Raviv, PhD
Florida Atlantic University, USA, mlevy2015@fau.edu, ravivd@fau.edu

***Abstract*– Given the mission critical nature of data centers, metrics must provide a holistic understanding of the data center behavior. After thoroughly reviewing existing data center metrics, this paper proposes a novel framework for more accurate metrics for data centers using a multidimensional approach. Performance is examined across four different sub-dimensions: productivity, efficiency, sustainability, and operations. Risk specifically associated with each of those sub-dimensions is contemplated. External risks are also considered independently of performance, namely location risk, global risks and site infrastructure risk.**

Keywords—Data center, metrics, performance, efficiency, risk.

I. Introduction

With today's significant and increasing reliance on digital information, data centers play a key role in making sure this information is securely stored and constantly available for users. Data centers are energy intensive complexes. With substantial expected growth in this sector, a need to make data centers more energy efficient has emerged, while guaranteeing their reliability and availability requirements. Efforts undertaken to pursue this goal include the creation of legislations, standards and best practices to follow when designing, building and operating a data center, as well as metrics to monitor its performance and find areas of improvement.

Being a new and growing field subject to fast changing technologies, current practices fail to consider important pieces of information. Existing metrics are very specific, and do not take into account the overall performance of the data center. Furthermore, there is an imminent need for metrics to incorporate an assessment of the risk that the data center is exposed to.

The proposed concept addresses the most recent U.S. Data Center Usage Report (June 2016) [1] which states the need for future research on performance metrics for data centers that better capture efficiency, to help better understand areas of improvement. The main motivation of this paper is to propose a new multidimensional approach for metrics incorporating productivity, efficiency, sustainability, and operations, as well as measurements of all the different risks associated to the data center.

The goal is to standardize a process so that it becomes a best practice to evaluate data centers. The expected long-term outcome of this proposal is to help in redefining standards so that important components (namely risk) that are currently being overlooked, better fit in a single picture. This paper adds technical and scientific support to existing theoretical and practical work that has mostly been carried out from outside academia.

II. Background

A data center can be defined as a facility with all the resources required for storage and processing of digital information, and its support areas. Data centers comprise the required infrastructure (e.g., power distribution, environmental control systems, telecommunications, security, fire protection, automation) and information technology equipment ("ITE" includes servers for processing, data storage, and network/communication equipment). The data center sector is very dynamic. Equipment can be upgraded frequently, new equipment may be added, obsolete equipment may be removed, and simultaneously old and new systems can be in use. In a data center, several threats can cause failures, including technical issues and human errors. The cost of downtime depends on the industry, reaching thousands of dollars per minute. The average cost associated with data center downtime for an unplanned outage is approximately \$ 9,000 per minute, an increase of about 60% from 2010 to 2016 [2].

According to standards, best practices, and user requirements, data center infrastructure must satisfy stringent technical requirements in order to guarantee reliability, availability, and security, as they have a direct impact on cost and efficiency. An infrastructure with high reliability and availability must have system redundancy, and for that reason, will be more expensive, and likely less efficient.

Data centers can consume 40 times more energy than conventional office buildings. Information technology equipment alone can use almost 100 Watts per square foot [3], and the high concentration of ITE in a data center results in higher power densities, compared to conventional buildings. In the United States, there are approximately 3 million data centers, representing one data center for every 100 people [4]. In 2014, data centers in the U.S. consumed around 70 billion kWh that is 1.8% of total electricity consumption. Data center electricity consumption increased around 4% from 2010-2014, a large shift from the approximate 24% increase from 2005-2010. Energy use is also expected to grow 4% from 2014-2020, and U.S. data centers are predicted to consume around 73 billion kWh in 2020 [1].

Environmental impact of data centers varies depending on the energy source used. The differences between the lowest and the highest greenhouse gas (GHG) emissions associated with each energy source could be around a factor of 200 [5]. A report from Global eSustainability Initiative estimates that for the period from 2002 to 2020, the emissions associated with the IT

Digital Object Identifier (DOI): http://dx.doi.org/10.18687/LACCEI2017.1.1.387
ISBN: 978-0-9993443-0-9
ISSN: 2414-6390

sector will grow by 180%, and data centers sector emissions will grow by 240%. This growth rate is much faster compared to the 30% increase in total emissions from all sources [6]. On the other hand, the IT sector has contributed to a reduction in emissions in other sectors. A report from the American Council for an Energy-Efficient Economy (ACEEE) shows that for each kWh consumed by the IT sector in U.S. other 10 kWh are saved in other sectors, due to the increase in economic productivity and energy efficiency [7].

Since the deregulation of telecommunications in the U.S. through the Telecommunications Act of 1996, a number of standards and best practices related to telecommunications, and more recently to the Data Center industry have emerged. Appropriate standards are the obvious direction to design or evaluate a data center, from national codes (required), local codes (required), and performance standards (optional). Organizations such as TIA (Telecommunication Industry Association), BICSI (Building Industry Consulting Service International Inc.), ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers), Uptime Institute, IEEE (Institute of Electrical and Electronics Engineers), NFPA (National Fire Protection Association), ISO (International Organization for Standardization), NIST (National Institute of Standards and Technology), JDCC (The Japan Data Center Council) and others, have contributed to the creation of standards, recommendations and best practices for the Data Center industry. Governments also impose regulatory standards on data centers depending on the nature of the business. Organizations such as The Green Grid (TGG), U.S. Department of Energy, U.S. Environmental Protection Agency (EPA), U.S. Green Building Council (USBGC), European Commission (EU Code of Conduct for Data Centres), ASHRAE, Japan's Green IT Promotion Council, and others have made significant contributions by generating white papers, reports, metrics, best practices and other documents related to data center efficiency.

Data centers standards and best practices evolve continually adapting to novel needs addressing new key issues and challenges. Due to its elevated cost and mission critical task, data centers must be designed and built to meet the minimum standard, which assures basic performance and efficiency. However, minimum standards do not translate into best practices when the goal is achieving the highest possible reliability with the highest energy efficiency. Optional standards and best practices have contributed to achieving this goal.

III. Data Center Metrics

Metrics can be defined as measures of quantitative assessment in order to compare or track efficiency, performance, progress or other parameters over time. Through different metrics data centers can be evaluated in comparison to the goals established or to similar data centers. Inconsistencies or variations in measurements can produce a false result for a metric, and for that reason it is very important to standardize metrics. Much of the current metrics, standards and legislations on data centers is focused towards energy efficiency, as this has proven to be a challenge given the rapid growth of the sector and its energy intensive nature. In 2016, ASHRAE introduced the standard 90.4 (Energy Standard for Data Centers). It establishes the minimum threshold for data center energy efficient design, construction, and operation and maintenance, and utilization of renewable energy resources [8].

Furthermore, there is significant concern among legislators about how efficiently an IT facility uses energy. The Energy Efficiency Improvement Act of 2014 (H.R. 2126) demands federal data centers to implement energy efficiency standards. This is motivated by the fact that federal data centers energy consumption represents 10% of all data centers in the U.S. This bill encourages federal data centers to improve energy efficiency and develop best practices to reduce energy consumption [9]. According to the Data Center Optimization Initiative (DCOI), federal data centers must reduce their power usage efficiency below a specified threshold by September 2018, unless they are scheduled to be shutdown, as part of the Federal Data Center Consolidation Initiative (FDCCI). These data centers must replace manual collection and reporting with automated infrastructure management tools by the end of 2018, and must address different metric targets including energy metering, power usage effectiveness, virtualization, server utilization and facility monitoring [10].

One of the most known energy efficiency metrics is Power Usage Effectiveness (PUE), introduced in 2007 by TGG [11]. Fig. 1 shows the energy flow in a data center. PUE is calculated as the ratio of total facility energy used to energy delivered to ITE [12]. In addition, Data Center infrastructure Efficiency (DCiE) is defined as the reciprocal of PUE [13]. Organizations such as the EPA have selected PUE as the metric to analyze energy performance in data centers, and define Source PUE as the ratio of total facility energy used to UPS energy [14], [15]. The main limitation of PUE is that it only measures the efficiency of the building infrastructure supporting a given data center, but it indicates nothing about the efficiency of ITE, or operational efficiency, or risk involved [1] [16][17].

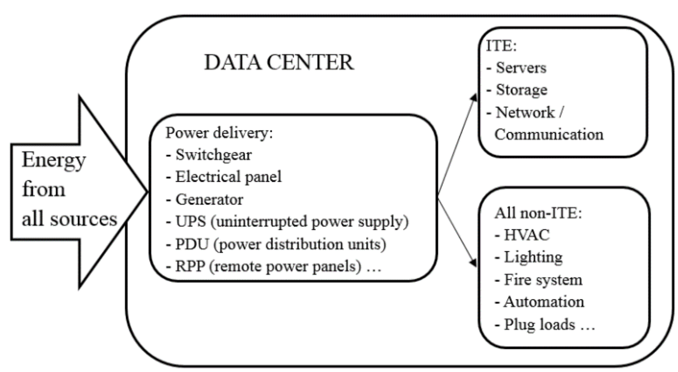

Fig. 1 Data center energy for PUE estimation

TGG has also introduced different metrics to measure -efficiency. Data Center Density (DCD) is defined as the total power consumption of all equipment divided by the area. Compute Power Efficiency (CPE) is estimated as the ITE utilization multiplied by the IT power consumption and divided by the total facility power consumption [18]. BCS Data Centre Specialist Group proposed fixed and proportional overhead metrics, to understand how energy and cost impacts the use of ITE. The fixed part of the energy consumption is considered if all ITE are not in use, and the proportional part takes into account the ITE load [19].

Different organizations have proposed metrics to measure server efficiency and performance for computers servers and storage. Standard Performance Evaluation Corporation (SPEC) has contributed with proposals such as SPEC Power and Performance Benchmark Methodology, which are techniques recommended for integrating performance and energy measures in a single benchmark; SPECpower_ssj2008, a general-purpose computer server energy-efficiency measure, or power versus utilization; SPECvirt_sc2013 for consolidation and virtualization; SPEComp2012 for highly parallel complex computer calculations; SPECweb2009 for web applications [20]. TheGreen500 ranks the most energy efficient supercomputers using the metric flops per Watt, to measure the performance per unit of power. Flops means floating-point operations per second, which represents the computer speed [21]. The Transaction Processing Performance Council (TPC) uses benchmarks with workloads that are specific to database and data management such as: TPC-Energy, which focuses on energy benchmarks (e.g. TPC-C and TPC-E for online transaction processing, TPC-H and TPC-DS for business intelligence or data warehouse applications, TPC-VMS for virtualized environment) [20]. VMWare proposed VMmark to measure energy consumption, performance and scalability of virtualization platforms [20]. The Storage Networking Industry Association (SNIA) Emerald Program and the Storage Performance Council (SPC) have contributed to establishing measurements for storage performance and the power consumption associated with the workloads [20]. Sun Microsystems introduced Space Wattage and Performance (SWaP) metric to measure server efficiency in a more comprehensive way. Space addresses the height in rack units (RUs). Wattage is the power consumption during actual benchmark runs or from technical documentation. Performance is measured by industry standard benchmarks (e.g., SPEC). In summary, SWaP = Performance/ (Space x Power) [22].

Considering the energy-proportional computing, where computing systems consume energy in proportion to the work performed, different metrics have been proposed. The idle-to-peak power ratio (IPR) metric is defined as the ratio system's idle consumption with no utilization over the full utilization power consumption. The linear deviation ratio (LDR) metric, shows how linear power is, compared to utilization curve. [23]

Metrics related to data center sustainability have also been introduced. Green Energy Coefficient (GEC) measures how much of the total energy is sourced from alternative energy providers such as solar, wind or geothermal plants, and encourage the use of renewable energy [24]. Carbon Usage Effectiveness (CUE) measures carbon emissions, total CO_2 emissions, in relation to ITE energy consumption [25]. Water Usage Effectiveness (WUE) measures total water usage in relation to ITE energy consumption [26]. In addition, the Energy Reuse Effectiveness (ERE) shows how energy is reused outside of the data center. It is calculated as the total energy minus reuse energy divided by ITE energy, or the Energy Reuse Factor (ERF) calculated as the reuse energy divided by the total energy [27].

The Uptime Institute has proposed different metrics to quantify energy consumption related to environmental sustainability, defining four categories of metrics: IT strategy, IT hardware asset utilization, IT energy and power efficient hardware deployment, and Site physical infrastructure overhead [28]. For those metrics different factors are defined, such as, the Site-Infrastructure Power Overhead Multiplier (SI-POM) estimated as the power consumption at the utility meter divided by the total power consumption at the plug of all ITE; the IT Hardware Power Overhead Multiplier (H-POM), estimated as the ratio of the AC hardware load at the plug and the DC hardware compute load, showing the ITE efficiency; the Deployed Hardware Utilization Ratio (DH-UR), estimated as the number of servers running live applications divided by the total servers deployed, or as the ratio of terabytes of storage holding data and the total of terabytes of storage deployed; the Deployed Hardware Utilization Efficiency (DH-UE) measured as the ratio of minimum numbers of servers required for peak compute load and the total number of servers deployed, which shows the possibilities of virtualization; and free cooling [28]. In addition, the Free Cooling metric to estimate potential savings using outside air; and the Energy Save metric to estimate the amount of money, energy or carbon emissions savings if the ITE hibernates while it is not used [28].

Different metrics have been developed to measure HVAC system performance. Rack Cooling Index (RCI) measures how efficient the cooling is for ITE cabinets compared to the ITE intake temperature [29]. Return Temperature Index (RTI) measures how the air management system performs [30]. HVAC effectiveness measures the ratio of ITE energy compared to the HVAC system energy. Airflow efficiency measures the total fan power needed by unit of airflow (Total fan power / Total fan airflow). Cooling system efficiency in terms of the average cooling system power divided by the average data center cooling load [15].

After recognizing the need for performance metrics that better capture the efficiency of a given data center, different entities have proposed metrics that measure the functionality of the data center (e.g. amount of computations it performs) and relate that to energy usage. An example of these are metrics to track useful work produced at a data center compared to power or energy consumed producing it. "Useful work" is defined as the tasks executed in a period of time, each one having a specific weight. These metrics are very specific to each user's activity but could provide a framework to develop comparisons between different data centers.

TGG proposed Data Center Performance Efficiency (DCPE) metric, defined as the ratio of useful work by total facility power [18]. In addition, TGG introduced the family Data Center Productivity (DCxP) metrics, such as Data Center energy Productivity (DCeP), measuring useful work produced divided by the total data center energy consumed producing this work [31]; Data Center compute Efficiency (DCcE) showing the efficiency of the compute resources, intended to find areas of improvement but not designed to compare different data centers [32]; and Data Center Storage Productivity (DCsP), defined as the ratio of useful storage system work to energy consumed [33].

The BCS Data Centre Specialist Group proposed the Data Center Fixed to Variable Energy Ratio (DC-FVER) metric, defined as the fixed energy divided by the variable energy plus one (1 + fixed energy / variable energy). It shows the inefficiencies through the wasted energy not delivering 'useful work'. The metric reflects the proportion of energy consumption that is variable, considering ITE, software and infrastructure [34]. Ebay released an initiative known as Digital Service Efficiency (DSE) metric, which shows the productivity and efficiency of the infrastructure through performance, cost, environmental impact and revenue. Performance is measured by transactions (buy or sell) per energy, per user and per server. Cost is measured by amount of money per energy, per transaction and per server. Environmental impact is estimated in metric tons of carbon dioxide per energy and per transaction. Revenue is estimated per transaction and per user [35]. Future Facilities proposed the availability, capacity and efficiency (ACE) performance assessment, which factors in the availability of ITE during failures, the physical capacity available, and how efficient the cool air delivery to ITE is [36].

The Uptime Institute proposed different metrics to measure efficiency and productivity. The Corporate Average Datacenter Efficiency (CADE), estimated as the ITE efficiency factored by the facility efficiency (CADE = ITE efficiency × facility efficiency = ITE asset utilization × ITE energy efficiency × Site asset utilization × Site energy utilization) [37]. The Data Center Energy Efficiency and Productivity (DC-EEP) index, which is the result of multiplying the IT Productivity per Embedded Watt (IT-PEW) and the Site Infrastructure Energy Efficiency ratio (SI-EER) [38].The IT organization is mainly responsible for the IT-PEW. The SI-EER is measured dividing the power required for the whole data center by the conditioned power delivered to the IT equipment.

The Green IT Promotion Council (Japan) introduced the Datacenter Performance per Energy (DPPE). It considers four sub-metrics: ITE Utilization (ITEU), ITE Energy Efficiency (ITEE), Power Usage Effectiveness (PUE) and Green Energy Coefficient (GEC). ITEU is the ratio of total measured power of ITE to total rated power of ITE, promoting efficient operation of ITE. ITEE is the ratio of total rated capacity of ITE and total rated energy consumption of ITE, promoting the procurement of efficient ITE. PUE is the total energy consumption of the data center divided by the total energy consumption of ITE, promoting energy saving of facilities. GEC is the green energy divided by the total energy consumption of the data center, promoting the use of green energy. Then the metric is defined as DPPE = ITEU × ITEE × 1/PUE × 1/ (1-GEC) [39].

Recently, in 2016, TGG introduced the new Performance Indicator (PI) metric, to visualize the data center cooling performance in terms of the balance of the following metrics: thermal conformance, thermal resilience and energy efficiency [40] [41]. IT thermal conformance indicates the proportion of ITE operating inside recommended inlet air temperatures ranges during normal operation. IT thermal resilience shows if there is any equipment at risk of overheating in case redundant cooling units are not operating due to a failure or planned maintenance. Energy efficiency is measured through the PUE ratio, and it indicates how the facility is operated compared to pre-established energy efficiency ratings.

Existing data center metrics reviewed in this paper are listed in Table A.1 (Appendix A), classified by type and main promoter. In addition, there has been academic research in specific data center metrics, such as a modified PUE metric using power demand [42], PUE for a CCHP Natural gas or Biogas Fuelled Architecture [43], PUE for application layers [44], performance metrics for communication systems [45], load dependent energy efficiency metrics [46], workload power efficiency metric [47].

These joint efforts have significantly improved efficiency on the data center infrastructure so that energy consumption has started to flatten out over time [1]. Current metrics however, fail to incorporate important aspects such as the risk involved, data center space types (e.g., hyperscale, service provider, internal, server room, server closet [1]), or tier levels of the infrastructure [48]. This being the case, comparisons between data center scores with the purpose of evaluating areas of improvement is not an easy task. Particularly, currently there is no metric that examines performance and risk simultaneously. A data center may have high performance indicators, with a high risk of failure. Research must therefore be refocused to incorporate risks, management and performance. Having this information may work as an early warning system, so that mitigation strategies and actions are undertaken due to the mission critical nature of data centers.

IV. MULTIDIMENSIONAL APPROACH

A new generation of metrics can help understand the performance of new and existing data centers and their associated risk. We propose a data center multidimensional scorecard comprised of performance and risk. Performance is examined across four different sub-dimensions: productivity, efficiency, sustainability, and operations. In addition, risk associated with each of those sub-dimensions is contemplated. External risks are also considered independently of performance, namely location risk, global risks and site infrastructure risk.

Is important to highlight that correlation can exist between the different elements of the scorecard. Fig. 2 shows a diagram of the proposed multidimensional metric. Although in reality all elements can be correlated, this proposal has been simplified by assuming there is no correlation between different performance sub-dimensions. Performance, defined by productivity, efficiency, sustainability and operations, and risk, would be measured or estimated through different mechanisms explained below.

Fig. 2 Data center multidimensional metric

1. *Productivity.*

Provides an indication of work accomplished. It will be measured through:

1.1. Useful work

This includes transactions, amount of information processed, units of production, etc., per cost of the data center. Cost includes capital expenditures (fixed assets and infrastructure equipment) and operating expenses (energy, human resources, maintenance, insurance, taxes, among other). When including monetary values and comparing these metrics across time, all future values of money need to be brought to present value so the comparison is consistent.

To obtain the data to estimate the work produced, a process needs to be established where the user defines 'useful work' and costs for the specific data center and the information is sent directly to this system. Once processes are in place, calculations of this metric can be performed automatically in real-time.

1.2. Downtime

Actual downtime, in terms of length, frequency, and recovery time. A separate measurement within this category will calculate the impact of downtime on productivity, measured as the work that was not carried out as well as other indirect tangible and intangible costs due to this failure.

To obtain the data a process needs to be established where downtime data (time, date and duration) is sent to this system. To calculate the impact on productivity a scale could be defined based on previous reports of data center outages costs [2].

2. *Efficiency.*

Measures the energy efficiency of:

2.1. Site infrastructure. It is calculated with the ratio of total energy used to energy delivered to ITE.
2.2. ITE: IT equipment efficiency, power versus utilization.
2.3. Productivity or useful work per energy consumption.
2.4. Physical space per energy consumption.

Energy consumption, and utilization data can be collected automatically from different equipment.

3. *Sustainability.*

Measures:

3.1. Green energy sources: calculating the ratio of green energy to total energy. The data is obtained automatically from real-time measurements.
3.2. How environmentally friendly processes, materials and components are. This data is collected by conducting an analysis or audit of processes. It must only be updated if and when a process changes.
3.3. Carbon footprint: Depending on the type of energy, assign different values of carbon footprint. Use existing and widely known methods to evaluate the greenhouse gas emissions. This value can be automated.

4. *Operations.*

Indicates how well managed the data center is. This would include an analysis of operations and maintenance processes such as site infrastructure, ITE, human resources training, security (electronic and physical), among other. Data is collected by conducting an audit and analysis of systems and processes. Data on frequency and quality of maintenance needs to be collected and recorded. Data on human resources training need to be collected and recorded. Data on electronic and physical security must be evaluated and assessed against a pre-established scale.

5. *Risk.*

Data center performance cannot be completely evaluated if the risks that may impact it are not considered. Optimization must involve risk, defined as potential threats that, if materialized, could negatively impact the performance of the data center. This metric intends not only to measure performance for each process area, but also to associate it with its level of risk. That way, the user may implement actions to achieve the optimum performance and later adjust that performance to a tolerable level of risk, which may again deviate the metrics from their optimum performance. In the long term, if variables remain unchanged, this model will lead to a stable equilibrium.

Risk of these sub-dimensions of performance, as well as the external risk which is independent of performance, are also measured through the use of metrics. They can be described as a causal system, where output depends on present and past inputs. An important strategy to reduce probability of failure is redundancy of resources but this component may affect efficiency, productivity, operations and cost.

5.1. Productivity risk.

This risk is assessed as the downtime probability of occurrence times its impact. The probability is estimated using present and past data of downtime. The impact would consider the cost of downtime.

5.2. Efficiency risk.

It is estimated with the ratio of processing utilization, ITE physical space utilization, and ITE energy utilization, to their respective total capacities. It also considers projected growth. When utilization is close to or at capacity, there is no room for growth, which means the risk that projections will not be met is high. This directly influences performance.

5.3. Sustainability risk.

The historic behavior of the different green energy sources, the percentage composition of each source, and their probability of failure, should be considered for estimating the associated risk.

5.4. Operations risk.

It is assessed by the operational risk, including organization and personnel qualifications, quality and frequency of maintenance, life-cycle planning, training and other. Analyzing historical data in order to estimate probability of failure due to improper operation in the areas identified, and its impact on performance.

5.5. External risk.

Performance risks are not the only risks involved in the data center. There are other major risk factors that are external to the actual operation of the data center that must be considered in this analysis. They include, but are not limited to:

5.5.1. Location risk.

The first author of this paper has previously developed a methodology that identifies potential threats and assigns a weight according to the relevance and impact associated to the data center location, to rank data centers according to their site risks [49]. Threats are classified as natural disasters (e.g., seismic activity, flooding, wind, ground stability), transportation and adjacent properties (e.g., public roads, underground transportation, air traffic, railways, marine), services (e.g., energy, telecommunications, water), and other (e.g., business incentives, real estate, population, insurance, taxes). In addition, the methodology helps quantify the probability of occurrence of each event, and estimates the potential consequences or impact of each event, in order to calculate risk of each and all threats.

5.5.2. Global risks.

These include factors such as variations in cost of energy, shortage of energy or water, potential security breaches, local and global regulations, new industry standards.

5.5.3. Site infrastructure risk.

This risk can be estimated by evaluating the levels of redundancy of ITE and site facilities. Redundancy can help mitigate possible downtime due to failures on equipment or maintenance.

The proposed metric can be summarized into a function, time dependent, to be further developed. For a specific time instant, regardless of the correlation between different parameters, the data center score can be defined as:

Data Center Score =
$= f(P_1, P_2, P_3, P_4, R_1, R_2, R_3, R_4, R_E, W_1, W_2, W_3, W_4)$
$= f(P_i, R_i, R_E, W_i)$ with $i=1,\ldots,4$

Where: P_i: Performance.
R_i: Risk of process.
R_E: Risk external to processes.
W_i: Weight of each category.

And sub-indexes: 1: Productivity.
2: Efficiency.
3: Sustainability.
4: Operations.

The data center scorecard involves both technical and non-technical aspects of the data center, as failures and risks may not only be due to technical issues but also to non-technical ones such as human error. Such metrics should measure parameters and processes. All required data must be inputted into the system. Inputs should be automatic when possible as the scorecard should receive information directly from the different systems that measure parameters or variables to be considered in these metrics. Depending on the final score calculated with the value of the dimensions, the scorecard would rank each data center on a scale to be defined, which allows for tangible comparisons between different data centers, or 'before and after' on the same data center. Furthermore, the proposed metric has the possibility to incorporate new performance and risk measurements in the future, preventing it from becoming outdated.

Equipment needs permanent monitoring and maintenance to assure proper and efficient performance. Obtaining real-time data is not a trivial task, especially in existing data centers, without adequate instrumentation to collect the data [11]. This underscores the need for new approaches to data center monitoring and management systems [50]. Real-time data collection is required for reliable metrics. Measuring some parameters in real-time is a challenge, if the related process cannot be easily automated. Improvements by some equipment manufacturers include the ability to directly access measurements of power, temperature, airflow, and resource utilization for each device. These measurements may include parameters such as the air inlet temperature, airflow, outlet temperature, power utilization, CPU utilization, memory utilization, and I/O utilization. Platform level telemetry can transform data center infrastructure management, allowing direct data access from the equipment processor [51]. It should be noted that gathering the right data and understanding its nature is more important than simply collecting more data [52].

V. Conclusions

Given the mission critical nature of data centers, metrics must provide a holistic understanding of the data center behavior. The proposed new metric, using a multidimensional approach, provides a comprehensive view of the data center, around which to create a strategy. The strategy should improve the overall metric of the performance and risk values combined, even when it does not mean attempting to achieve the optimal performance. It is different from existing metrics in that it does not limit itself to measuring a value whose optimization can be automated, instead, this is only part of its scope. In summary, the new metric should be capable of standardizing a process so that it becomes a best practice to help evaluate data centers and compare them to each other. The metric can provide a ranking of how the data center operates, and allows comparison between scenarios. The outcome is a scorecard, considering productivity, efficiency, sustainability, operations, and risk, that assists in finding areas of improvement, which should be strategically addressed, and will also constitute a strong basis for decision-making.

VI. Future Research

A preliminary approach on data center performance and risk metrics is presented in this paper, however, further research needs to be conducted to find adequate multidimensional metrics and make a decision on which ones are most reasonable to use for the intended purpose.

Since there are currently no solutions available that include all proposed factors, a simulation software can be developed to validate and calibrate the metric. There are some initiatives using an open source strategy that can be used as the starting point for the development: the OpenDCIM infrastructure management tool [53], the OpenFoam CFD toolbox [54], the CloudSim cloud computing simulator [55], the Greencloud cloud computing simulator [56], and the iCanCloud simulator [57]. None of these tools has the ability to express operational and design risk, nor do they integrate specific component models.

Different data center control strategies can be deployed to improve metrics, for which a theoretical study of possible strategies must be conducted. Many of them could require simulation tools to validate them before implementation. By tracking the proposed metric, the data center personnel will have an overall understanding of the data center behavior, including performance and risk in a multidimensional view. In addition, the multidimensional metric could be transformed through different operators as a composite metric with just one value.

Actions undertaken will impact the metric in real time, and the system must be able to recognize it. When issues identified by each of these metrics are addressed, processes can be optimized or moved closer to their optimal point based on the vision for the specific data center. That way, when variables are re-measured, the result of the metrics would have improved. To recognize trends or predict future behavior, tools such as predictive analysis and machine learning are required. Machine learning considers algorithms that can learn from and make predictions on data, rather than strictly static program code.

With all data in hand, and studying the correlation between the different metric scores and their associated risks, simulations of different scenarios can be run to visualize how a change in parameters impacts risk, and likewise, how a change in risk factors affects metric results.

References

[1] Ernest Orlando Lawrence Berkeley National Laboratory, “United States Data Center Energy Usage Report. LBNL-1005775,” 2016.

[2] Ponemon Institute LLC, “Cost of Data Center Outages. Data Center Performance Benchmark Series,” 2016.

[3] S. Greenberg, E. Mills, and B. Tschudi, “Best Practices for Data Centers : Lessons Learned from Benchmarking 22 Data Centers,” *2006 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA, pp. 76–87, 2006.

[4] U.S. Department of Energy. Office of Energy Efficiency & Renewable Energy, “10 Facts to Know About Data Centers,” 2014. [Online]. Available: http://energy.gov/eere/articles/10-facts-know-about-data-centers.

[5] I. Ahmad and S. Ranka, “Chapter 43: Overview of Data Centers Energy Efficiency Evolution,” in *Handbook of Energy-Aware and Green Computing*, vol. 2, CRC Press, 2012, pp. 983–1027.

[6] The Climate Group - Global eSustainability Initiative (GeSI), “SMART 2020: Enabling the low carbon economy in the information age,” 2008.

[7] J. A. Laitner and K. Ehrhardt-Martinez, “Information and Communication Technologies: The Power of Productivity,” American Council for an Energy-Efficient Economy, Washington, DC, 2008.

[8] ASHRAE, “Energy Standard for Data Centers,” ANSI/ASHRAE Standard 90.4, July 27, 2016.

[9] B. Howard, “H.R. 2126: The Energy Efficiency Improvement Act,” 2014. [Online]. Available: http://www.usgbc.org/articles/hr-2126-energy-efficiency-improvement-act.

[10] T. Scott, “Data Center Optimization Initiative (DCOI). Memorandum,” 2016.

[11] The Green Grid, “PUE™: A Comprehensive Examination of the Metric. White paper # 49,” 2012.

[12] The Green Grid, “Green Grid Metrics: Describing Data Center Power Efficiency. Technical Committee White Paper,” 2007.

[13] The Green Grid, “Green Grid Data Center Power Efficiency Metrics: PUE and DCIE. White paper # 6,” 2008.

[14] U.S. Environmental Protection Agency, “Energy Star for Data Centers Scheduled Portfolio Manager Release on June 7 , 2010,” 2010.

[15] W. Lintner, B. Tschudi, and O. VanGeet, “Best Practices Guide for Energy-Efficient Data Center Design,” U.S. Department of Energy, 2011.

[16] N. Horner and I. Azevedo, “Power usage effectiveness in data centers: overloaded and underachieving,” *The Electricity Journal*, vol. 29, no. 4. Elsevier Inc., pp. 61–69, 2016.

[17] V. Avelar, D. Azevedo, and A. French, *PUE™: A Comprehensive Examination of the Metric*. Atlanta, GA: ASHRAE Datacom Series, 2012.

[18] P. P. Ray, “The green grid saga - a green initiative to data centers: a review,” *Indian Journal of Computer Science and Engineering*, vol. 1, no. 4. pp. 333–339, 2010.

[19] L. Newcombe, “Data centre energy efficiency metrics: Existing and proposed metrics to provide effective understanding and reporting of data centre energy. White Paper,” BCS Data Centre Specialist Group, 2010.

[20] K.-D. Lange and K. Huppler, *Server Efficiency: Metrics for Computer Servers and Storage*. Atlanta, GA: ASHRAE Datacom Series, 2015.

[21] S. Sharma, C. H. Hsu, and W. C. Feng, "Making a Case for a Green500 List," *20th IEEE International Parallel & Distributed Processing Symposium (IPDPS) Workshop on High-Performance, Power-Aware Computing (HP-PAC)*. Rhodes Island, Greece, 2006.

[22] Sun Microsystems Inc, "Sun Sparc ® Enterprise T5120 and T5220 Server Architecture. White Paper," 2007.

[23] I. Ahmad and S. Ranka, "Chapter 44: Evaluating Performance, Power, and Cooling in High-Performance Computing (HPC) Data Centers," in *Handbook of Energy-Aware and Green Computing*, vol. 2, CRC Press, 2012, pp. 1029–1049.

[24] The Green Grid et al., "Harmonzing Global Metrics for Data Centers Energy Efficiency," 2012.

[25] The Green Grid, "Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric. White paper # 32," 2010.

[26] The Green Grid, "Water Usage Effectiveness (WUE): A Green Grid Data Center Sustainability Metric. White paper # 35," 2011.

[27] The Green Grid, "ERE: A Metric for Measuring the Benefit of Reuse Energy From a Data Center. White paper # 29," 2010.

[28] J. R. Stanley, K. G. Brill, and J. Koomey, "Four Metrics Define Data Center Greenness," Uptime Institute, 2007.

[29] M. K. Herrlin, "Rack Cooling Effectiveness in Data Centers and Telecom Central Offices: The Rack Cooling Index (RCI)," in *ASHRAE Transactions*, vol. 111, no. 2, Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2005, pp. 725–731.

[30] M. K. Herrlin, "Airflow and Cooling Performance of Data Centers: Two Performance Metrics," in *ASHRAE Transactions*, vol. 114, no. 2, Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2008, pp. 182–187.

[31] The Green Grid, "A framework for data center energy productivity. White paper # 13," 2008.

[32] The Green Grid, "The green grid data center compute efficiency metric: DCcE. White paper # 34," 2010.

[33] The Green Grid, "The Green Grid Data Center Storage Productivity Metrics (DCsP): Application of Storage System Productivity Operational Metrics. White paper #58," 2014.

[34] L. Newcombe, Z. Limbuwala, P. Latham, and V. Smith, "Data centre Fixed to Variable Energy Ratio metric DC - FVER: An alternative to useful work metrics which focuses operators on eliminating fixed energy consumption. White Paper," BCS Data Centre Specialist Group, 2012.

[35] eBay Inc, "Digital Service Efficiency - Solution Brief. White paper," San Jose, CA, 2013.

[36] D. King and S. Davies, "ACE Performance Assessment - Empowering the Data Center Operator. White paper," Future Facilities, 2014.

[37] J. Kaplan, W. Forrest, and N. Kindler, "Revolutionizing data center energy efficiency." McKinsey & Company, pp. 1–13, 2008.

[38] K. G. Brill, "Data Center Energy Efficiency and Productivity," Uptime Institute, 2007.

[39] Green IT Promotion Council, "Concept of New Metrics for Data Center Energy Efficiency - Introduction of Datacenter Performance per Energy [DPPE]," Japan, 2010.

[40] The Green Grid, "New Metric To Assess and Visualize Data Center Cooling," 2016. [Online]. Available: http://www.thegreengrid.org/~/media/Press Releases/The Green Grid_Press Release_Performance Indicator New metric_vFINAL.pdf.

[41] The Green Grid, "The Performance Indicator Assessing and Visualizing Data Center Cooling Performance. White Paper," 2016.

[42] E. Jaureguialzo, "PUE: The Green Grid metric for evaluating the energy efficiency in DC (Data Center). Measurement method using the power demand," *2011 IEEE 33rd International Telecommunications Energy Conference (INTELEC)*. Amsterdam, 2011.

[43] F. De Angelis and U. Grasselli, "The next generation green data center a modified power usage effectiveness metric proposal for a CCHP natural gas or biogas fuelled architecture," *2015 IEEE 15th International Conference on Environment and Electrical Engineering, (EEEIC)*. Rome, pp. 907–911, 2015.

[44] R. Zhou, Y. Shi, and C. Zhu, "AxPUE: Application level metrics for power usage effectiveness in data centers," *2013 IEEE International Conference on Big Data*. Silicon Valley, CA, pp. 110–117, 2013.

[45] C. Fiandrino, D. Kliazovich, P. Bouvry, and A. Y. Zomaya, "Performance Metrics for Data Center Communication Systems," *2015 IEEE 8th International Conference on Cloud Computing*. New York, NY, pp. 98–105, 2015.

[46] D. Schlitt and W. Nebel, "Load dependent data center energy efficiency metric based on component models," *2012 International Conference on Energy Aware Computing*. Guzelyurt, Cyprus, 2012.

[47] T. Wilde *et al.*, "DWPE, a new data center energy-efficiency metric bridging the gap between infrastructure and workload," *2014 International Conference on High Performance Computing & Simulation (HPCS)*. Bologna, pp. 893–901, 2014.

[48] P. Turner, J. H. Seader, V. Renaud, and K. G. Brill, "Tier Classifications Define Site Infrastructure Performance," Uptime Institute, 2006.

[49] M. Levy, "Análisis de Riesgos y Selección de Localidad: Dos Pilares Imprescindibles [Risk analysis and site selection: two essential pillars]," *Datacenter Dynamics Focus*, pp. 32–34, 2014.

[50] M. Levy and J. O. Hallstrom, "A New Approach to Data Center Infrastructure Monitoring and Management (DCIMM)," *IEEE CCWC 2017. The 7th IEEE Annual Computing and Communication Workshop and Conference*. Las Vegas, NV, 2017.

[51] J. L. Vincent and J. Kuzma, "Using Platform Level Telemetry to Reduce Power Consumption in a Datacenter," *2015 31st Thermal Measurement, Modeling & Management Symposium (SEMI-THERM)*. Folsom, CA, 2015.

[52] J. Koomey and J. Stanley, "The Science of Measurement: Improving Data Center Performance with Continuous Monitoring and Measurement of Site Infrastructure," 2009.

[53] OpenDCIM Foundation Inc, "The Data Center Infrastructure Management open source application," 2014. [Online]. Available: http://opendcim.org/.

[54] OpenFoam, "open source CFD toolbox." [Online]. Available: www.openfoam.com.

[55] The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, "CloudSim: A Framework For Modeling And Simulation Of Cloud Computing Infrastructures And Services." [Online]. Available: http://www.cloudbus.org/cloudsim/.

[56] Universite Du Luxembourg, "Greencloud - The green cloud simulator." [Online]. Available: https://greencloud.gforge.uni.lu/.

[57] A. Nuñez, J. L. Vazquez-Poletti, A. C. Caminero, G. G. Castañe, J. Carretero, and I. M. Llorente, "iCanCloud: A flexible and scalable cloud infrastructure simulator," *Journal of Grid Computing*, vol. 10, no. 1. pp. 185–209, 2012.

APPENDIX A

The following table presents a summary of the reviewed existing data center metrics, classified by type with the main promoter. In some cases the main promoter could not be identified. Must be noted that each metric uses its own definition for some terms (e.g., efficiency, productivity, performance) and for comparison purposes the same metric must be used.

Table A.1. Existing data center metrics

Metric	Type	Promoter
Power Usage Effectiveness (PUE) Data Center infrastructure Efficiency (DCiE)	Energy efficiency.	The Green Grid
Data Center Density (DCD)	Space efficiency	The Green Grid
Fixed and proportional overhead	Energy efficiency.	BCS Data Centre Specialist Group
Compute Power Efficiency (CPE)	ITE (server) efficiency	The Green Grid
SPECpower_ssj2008	ITE (server) efficiency	Standard Performance Evaluation Corporation (SPEC)
SPECvirt_sc2013	ITE (server) efficiency. Consolidation and virtualization.	SPEC
SPEComp2012	ITE (server) efficiency. Highly parallel complex computer calculations.	SPEC
SPECweb2009	ITE (server) efficiency. Web applications.	SPEC
TheGreen500	ITE (server) efficiency	TheGreen500
TPC-Energy: TPC-C and TPC-E	ITE (server) efficiency. Online transaction processing	Transaction Processing Performance Council (TPC)
TPC-Energy: TPC-H and TPC-DS	ITE (server) efficiency. Business intelligence or data warehouse applications	TPC
TPC-Energy: TPC-VMS	ITE (server) efficiency. Virtualized environment	TPC
Vmmark	ITE (server) efficiency. Virtualization platforms	VMWare
Storage performance	ITE (storage) efficiency.	Storage Networking Industry Association (SNIA) Emerald Program and the Storage Performance Council (SPC)
Space Wattage and Performance (SWaP)	ITE (server) efficiency.	Sun Microsystems
Idle-to-peak power ratio (IPR)	ITE (server) efficiency. Idle consumption.	---
Linear deviation ratio (LDR)	ITE (server) efficiency. Power linearity.	---
Green Energy Coefficient (GEC)	Sustainability. Alternative energy.	The Green Grid
Carbon Usage Effectiveness (CUE)	Sustainability. Carbon emissions.	The Green Grid
Water Usage Effectiveness (WUE)	Sustainability. Water usage.	The Green Grid
Energy Reuse Effectiveness (ERE) Energy Reuse Factor (ERF)	Sustainability. Energy reuse.	The Green Grid
Site-Infrastructure Power Overhead Multiplier (SI-POM)	Sustainability. Site physical infrastructure overhead.	Uptime Institute
IT Hardware Power Overhead Multiplier (H-POM)	Sustainability. ITE efficiency.	Uptime Institute
Deployed Hardware Utilization Ratio (DH-UR)	Sustainability. ITE efficiency.	Uptime Institute
Deployed Hardware Utilization Efficiency (DH-UE)	Sustainability. ITE efficiency.	Uptime Institute
Free Cooling	Sustainability. Free cooling.	Uptime Institute
Energy Save	Sustainability. ITE hibernate.	Uptime Institute
Rack Cooling Index (RCI). Return Temperature Index (RTI). HVAC effectiveness. Airflow efficiency. Cooling system efficiency.	Performance. Cooling system.	American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). ANCIS Inc.
Data Center Performance Efficiency (DCPE)	Performance. Useful work.	The Green Grid
Data Center energy Productivity (DCeP)	Performance. Useful work.	The Green Grid
Data Center compute Efficiency (DCcE)	Performance. Compute resources.	The Green Grid
Data Center Storage Productivity (DCsP)	Performance. Storage systems.	The Green Grid
Data Center Fixed to Variable Energy Ratio (DC-FVER)	Performance. Wasted energy. Useful work.	BCS Data Centre Specialist Group
Digital Service Efficiency (DSE)	Performance, cost and sustainability.	Ebay
Availability, Capacity and Efficiency (ACE)	Performance and efficiency (cooling system).	Future Facilities
Corporate Average Datacenter Efficiency (CADE)	Performance. Efficiency.	Uptime Institute
Data Center Energy Efficiency and Productivity (DC-EEP)	Performance. Efficiency and productivity	Uptime Institute
Datacenter Performance per Energy (DPPE)	Performance.	The Green IT Promotion Council (Japan)
Performance Indicator (PI)	Performance. Cooling system.	The Green Grid