Mathematical Biology and Bioinformatics
2023. V. 18. № 1. P. 215–227. doi: 10.17537/2023.18.215

========== INFORMATION AND COMPUTER ==========
========== TECHNOLOGIES IN BIOLOGY AND MEDICINE ==========

Data Center Efficiency Model: A New Approach and the Role of Artificial Intelligence

Isaev E.A.[1], Kornilov V.V.[2], Grigoriev A.A.[3]

[1]*Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Moscow, Russia*
[2]*National Research University Higher School of Economics, Moscow, Russia*
[3]*Financial University under the Government of the Russian Federation, Moscow, Russia*

Abstract. Bioinformatics technologies play a significant and growing role in life science research, and as these technologies develop, so does the complexity of data. The challenge of biological data growth has given rise to a number of bioinformatics data centers that offer services and solutions ranging from large-scale biosystems analyze that accounts for entire OMICs to nanoscale experiments where molecular modeling can provide insight o structure and dynamics of molecular complexes of biological components. Obviously, this kind of research requires a highly specialized level of computational and statistical expertise, as well as high-performance resources. The importance of information technology is growing, as is the use of computer information systems throughout the world. There are more and more specialized data centers and they consume more energy. The development of new strategies for energy efficiency of data centers is becoming relevant. These strategies aim to reduce the amount of energy consumed by data centers and their environmental impact without sacrificing performance. The article examines performance metrics, proposes a new method for data center energy efficiency, and discusses the role of artificial intelligence techniques in achieving these goals.

Key words: *data centers; green data centers; energy efficiency; artificial intelligence.*

1. INTRODUCTION

A data center is a specialized technical complex that houses computing systems, data storage systems, telecommunications equipment, and engineering infrastructure to provide the IT equipment with the conditions it needs to operate. This includes providing an uninterrupted and guaranteed power supply, maintaining the requisite climatic conditions for operation, and implementing the necessary security measures continuously [1]. This means that data centers are widely recognized as a standard method of providing users with access to computing resources because they provide a secure technological environment for data processing, storage, and transmission [2]. The total cost of IT infrastructure can be cut by optimizing the cost of technical facilities, administration, and operation through data center consolidation [3, 4].

The IT industry is expanding at a rapid rate, society is becoming more and more informed, the fourth industrial revolution is developing, and businesses are becoming more digital [5, 6]. This includes the shift caused by the widespread adoption of remote work as a result of the COVID-19 pandemic, which raises the demand for data center services, especially from business users [7]. The main trend is that more and more is being asked of the performance and reliability of the IT infrastructure, while the amount of corporate information being processed keeps going up. Meanwhile, people are asking for the cost of maintaining and developing the IT infrastructure to go down, so that IT infrastructure investment can be kept [8, 9]. Hence, there needs to be a good balance between the data center's need for high computing power and data capacity, its maintenance costs, and how quickly high-performance applications can be deployed.

It is not unusual for large and sophisticated data centers to house hundreds of thousands of physical servers. For example, the data center of the Chinese company Range International Information Group, which is regarded as one of the largest in the world, takes up a total area of 600 thousand square meters, which is equivalent to the space needed to accommodate 110 football fields [10]. The concept of Network-Critical Physical Infrastructure (NPCI) is applied when creating the technical infrastructure of the data center [11]. These are the most important components of an engineering infrastructure:

- power supply,
- air conditioning,
- physical placement of data center components,
- physical security (video surveillance systems, access control),
- fire protection,
- cable infrastructure.

In the meantime, powering engineering and computing systems account for the majority of operating costs associated with running a data center (up to 60 %), with air conditioning systems using nearly half of the electricity consumed [12]. Approximately 200 terawatts are the total power capacity of all data centers worldwide, easily exceeding the power needs of some small nations, and this power consumption does not stop rising. Data center operators have legitimate concerns as a result of the enormous energy requirements of data centers, which present growing challenges in terms of cost and sustainability [13, 14]. Even a small reduction in electricity consumption can result in significant savings. In addition to this, having a lower overall power consumption is better for the environment. If data centers could reduce their electricity use by some 1 %, they would save enough carbon dioxide to extend the lives of millions of trees [15]. Manufacturers of the engineering infrastructure for data centers are concentrating on using cutting-edge energy-saving technologies to achieve high levels of energy efficiency, but overall, this problem is still very crucial [16, 17].

This paper explores how artificial intelligence (AI) technologies contribute to data center efficiency and presents a novel method for assessing data center energy efficiency by focusing on the metrics that are critical to data center efficiency.

2. GREEN DATA CENTERS

The desire to stop the unchecked expansion of environmental resources is part of the global trend toward intensifying the fight against global climate change and the "sustainable development" of society. When viewed in this light, the problem of environmentally responsible electricity consumption, which is directly connected to the issue of thermal pollution in the environment, takes on a greater degree of significance. Not only does the prosperity of any modern nation's digital economy depend on the performance of its data centers in terms of efficiency, dependability, and security, but so does the health of the world's ecosystems as a whole. A green data center is intended to have the least possible environmental impact by maximizing energy efficiency [18]. The infrastructure and operations of green data centers are identical to those of conventional data centers, with the added benefit of being able to significantly reduce both energy use and carbon emissions [19, 20]. The paper [21] states that the energy efficiency and carbon emission level of IT equipment, including servers, storage devices, communication devices, and infrastructure (ventilation and cooling systems, blowers and power distributors), are the main determinants of the amount of greenhouse gases emitted by data centers. Market trends predicted that the global green data center market would reach $59.32 billion in value in 2021 and would then expand on average by 23.5 % annually through 2026 [22]. Green data centers can reduce the environmental impact of computing equipment by implementing innovative, energy-efficient solutions in data center infrastructure, such as new types of cooling systems [23], specifically direct liquid cooling technology [24, 25], the use of a new generation of lithium-ion batteries in data center power supply systems [26, 27], or the use

of renewable energy sources [28, 29]. Despite the obvious disadvantages of such solutions, such as the geographically limited number of possible deployment sites for such complexes, sustainable and renewable energy sources have become a trend in the development of green data centers. For example, the paper [30] advocates for the use of alternative energy sources, particularly solar and tidal energy. Indeed, this resource is constant and predictable, making tidal current kinetic energy an extremely competitive energy resource when compared to other widely used renewable energy sources. This paper discusses the rated performance of a hybrid tidal photovoltaic system to power a megawatt green data center and presents a mathematical model of the system under various operating conditions.

Increasing the efficiency of cooling systems is one way to drastically reduce data center energy consumption. The review [31] discusses cooling systems that incorporate thermal energy storage (TES) technology, as well as examples of how TES technology can be used in data centers, how to operate them, and how to meet high reliability and security requirements for data center power supply. Although TES-based electronics cooling and thermal energy storage are said to have the greatest potential for use in data centers to improve system energy efficiency performance, there are significant differences in the layout and operating conditions of data center cooling systems, resulting in a wide range of applications and TES performance. Therefore, TES integrated cooling systems require further improvements before they can be broadly used for the public market.

The development of cost-free cooling systems is still ongoing. These function on the principle that equipment is cooled by drawing in outside air (normal outside air). As the air moves through the filters and into the machine hall, it becomes purified and free of dust particles. This is the most energy-efficient cooling system available right now. Unfortunately, for obvious reasons, the method's effectiveness is immediately dependent on the outdoor air temperature. When the air temperature is high enough (above 25–27 °C), more technological solutions are required. One example of this would be adiabatic cooling of the outdoor air, which is based on the physical effect of lowering the temperature of dry air while simultaneously increasing the air's humidity (passing through a wet surface). Urban areas with average humidity levels of 20 % or less can benefit greatly from this system. The authors [32] consider the possibility of using solar chimney technology for air cooling in green data centers. This is a passive heating and cooling system that uses solar energy and natural ventilation to regulate the temperature inside buildings. This study proposes an innovative theoretical model of a solar chimney-based direct air cooling system for estimating the data center's internal thermal environment, ventilation flow rate, and heat recovery under given climatic conditions. The results show that the data center's thermal environment can meet cooling requirements under favorable climatic conditions, and that such a project has a high return on investment. The proposed cooling system is not only practical and affordable, but it is also an efficient way to reduce energy consumption and operating costs.

3. ARTIFICIAL INTELLIGENCE TECHNIQUES TO OPTIMIZE DATA CENTER ENERGY CONSUMPTION

The application of artificial intelligence in data centers is one example of a widespread movement toward the intelligent advancement of all digital technologies. The global AI market is expected to nearly tenfold to $299.64 billion by 2026, at a compound annual growth rate of 35.6 %, according to the Facts and Factors report [33]. Notably, Google, Microsoft, and IBM, along with some other companies, are the key players in this market. This is because they are also among the largest providers of cloud services based on the size of their data centers. More than 30 % of data centers without AI and machine learning could soon become less competitive from an operational and financial standpoint, according to analyst firm Gartner's report from 2019 [34]. The use of AI applications is essential for optimizing the tasks of any data-driven business, and is unquestionably essential for automating data center operations that are initially

data-driven. The design phase, data center capacity planning, the operational phase, and data center optimization are some of the stages of data center creation and operation where AI applications are starting to be used. To ensure that enormous amounts of data are processed efficiently, for instance, workload balancing of servers is necessary. In this situation, AI applications can provide effective workload distribution between servers using predictive analytics, learning from the analysis of data accumulated during data center operations. Predicting failures and minimizing downtime are important challenges related to data center optimization. When large amounts of data are collected, analyzed, and processed using machine-learning models, it is possible to make accurate predictions regarding potential issues and equipment failures, as well as optimize maintenance schedules and other related matters. Another task that AI algorithms can help with is improving data center security. AI can spot cyber threats, find security gaps in data centers, examine incoming and outgoing data for security threats, and detect malware by observing typical network behavior and examining deviations from it. Physical and data center security systems can both benefit from AI. It can recognize faces, analyze employee behavior patterns, compare those patterns to available data, detect abnormal behavior, and alert security personnel.

When it comes to building and running data centers, another important step is to find out how to optimize energy use and make the best use of energy resources based on AI systems. Data centers manage applications that rely heavily on storage and I/O, memory, networking, and other resources. When one type of application uses too much of the processor, disk, or network, it can slow down the system and use a lot more energy.

Installing a data center sensor system in data centers is one way to address this issue. Indeed, this enables an intelligent control system to not only monitor all characteristics, including hardware and air temperature, energy consumption, computing equipment load, etc., but also to identify patterns that can be used to increase energy efficiency by examining how many variables interact [35].

Another solution is to use virtualization technology, which allows multiple applications and operating systems to run on fewer servers, allowing green data centers to be built. Standard methods of virtual machine deployment, on the other hand, frequently cause physical servers to be used less often, which wastes a lot of power in the data center. This hurts service providers and makes it harder for potential customers to use their services. The solution is to find ways and algorithms to combine tasks by moving them to virtual machines. This will make better use of physical compute servers and use less power in the data center [36–41]. The placement of virtual machines is a crucial strategy for reducing data center energy consumption in these and other related studies. The emphasis is on efficient virtual machine placement on servers to optimize physical resources used (memory, bandwidth, processor, etc.), network resources used, or cooling energy consumption. The methods presented can optimize data center energy consumption based on one, two, or, less frequently, three factors (e.g. server, network, or cooling) [42]. However, as more resource factors are taken into account for the relocation of virtual machines to be optimized, the problem's complexity and the scope of applications for algorithmic solutions both dramatically increase [43]. To solve this problem, new algorithms employ advanced AI techniques. To set a dynamic hybrid resource deployment rule, for example, the authors [44] could modify the *K*-means clustering algorithm for unsupervised learning and the KNN classification algorithm for teacher-assisted learning. Next, based on machine learning theory, a dynamic hybrid machine learning-based energy information resource deployment algorithm for cloud data centers is proposed.

Intelligent Data Center Infrastructure Management (DCIM) systems are another popular way to improve data center efficiency. These systems provide managers with greater control over the operation of all data center facilities. Administrators can plan and manage equipment operations more effectively, identify potential risks, prevent breakdowns, and thus reduce downtime. Individual server underutilization is a major contributor to data center inefficiencies. DCIM

increases efficiency by allowing administrators to identify which servers are not performing useful tasks and either shut them down or overload them. Additionally, the system can accurately estimate how other devices are being used and calculate the data center's overall energy consumption.

Here are a few recent instances of successful applications of AI to lower data center power usage and carbon footprint.

By market share, Alibaba Cloud is the third largest public cloud service provider in the world. It launched a data center in Germany that was tailored to accommodate applications involving AI and machine learning. The new facility is located in Frankfurt and serves clients from the manufacturing, retail, and automotive industries. This data center has a dry cooler instead of a mechanical cooler. It will run on 100 % green electricity and use a cloud-based platform to track and optimize that company's daily carbon footprint [45].

Back in 2016, Google reported that it had successfully used DeepMind AI technology to optimize and fully automate the cooling system in its data centers [46]. In response to external and internal factors, data center systems are automatically controlled in real-time. More than 120 parameters, including air conditioning control, window closing and opening, fan speed, and others, are monitored by the intelligent control system. Depending on the load on the server equipment, the control algorithms determine which cooling system configurations will reduce energy consumption and automatically turn them on. They do this by using an effective predictive model of energy consumption. This caused the cooling system to use 40 % less energy. By the year 2030, all of Google's data centers will be powered by carbon-free energy round-the-clock, as part of the company's pledge to decarbonize its energy consumption in its entirety. Some Google data centers already use 90 % carbon-free energy; the overall average was 61 % in 2019 and 67 % in 2020. The company's best practices for energy and carbon footprint will reduce energy consumption by a factor of 100 and emissions by a factor of 1,000 [47].

4. DATA CENTER ENERGY EFFICIENCY ASSESSMENT MODEL

A thorough analysis of energy efficiency is the first step toward achieving a more energy-efficient data center. Power Usage Efficiency (PUE) is the ratio of the total amount of power used by the IT equipment in a data center to the total amount of power used by the infrastructure of the data center. The PUE was developed by a working group of government agencies and industry leaders as part of the Green Grid consortium, which focuses on energy efficiency and emission reduction issues in data centers. PUE is now the de facto industry standard for measuring energy efficiency. Alongside this, Green Grid has developed and supports two additional metrics to ensure that data centers are more accurately described. Two such metrics exist, Carbon Usage Effectiveness (CUE) and Water Usage Effectiveness (WUE), and they measure the efficiency with which carbon dioxide is released into the atmosphere and water is used in data centers, respectively. To make index computation more manageable during data center operations, the index-specific definition methodology and the software module that accompanies it were both developed. The CUE index was introduced with the intent of accurately displaying and tracking the amount of carbon dioxide that the data center generated and released into the atmosphere over time:

$$\text{CUE} = \text{Carbon Usage} / \text{IT Equipment Energy}.$$

The WUE index should show how much water is used to cool the data center. The air-conditioning capacity is known to be divided into the apparent and latent ones, and only the apparent part is used for real air cooling. The latent cooling capacity is mostly wasted as condensate, which is then drained off and sent to the sewage system. Furthermore, precision air conditioners are outfitted with steam-humidifiers, the power consumption of which exceeds that of the air conditioners (compressor and fans) to ensure that humidity in data centers is restored. Thus, dehumidification of air in data centers due to inefficient climatic equipment settings results

in both a loss of cooling capacity and additional energy costs for the humidification system. The WUE index aims to quantify such costs:

WUE = Water Usage / IT Equipment Energy.

The CUE and WUE indices have ideal values of zero (unlike PUE, which has a minimum value of one). The PUE, CUE, and WUE metrics, according to The Green Grid, assist customers in selecting the most efficient data center, as well as the operations department in selecting the most efficient mode of data center operation.

However, while the approach outlined for evaluating data center energy efficiency is popular, it is not widely accepted; in fact, many data centers have yet to implement a system of evaluation metrics and, as a result, have not developed energy-saving standards.

As part of this work, it is suggested that a systematic model of data center efficiency accounting be built, taking into account some different metrics.

Let us use two parameters to assess data center energy efficiency: the traditional PUE and the Data Center Infrastructure Efficiency (DCiE) ratio. The DCiE is the inverse of PUE and displays the percentage of the facility's total energy consumption that is consumed by the IT load.

Data center energy efficiency in monetary terms can be calculated using the formula below:

$$E = N_t \cdot \left(\frac{C_e \cdot E\%}{100\%} \right),$$

where N_t is the number of computing tasks (useful work of the data center);
C_e is the cost of electricity;
E% is the total efficiency of the new technology implemented.

The efficiency of the new technology implemented is a comprehensive measure of the data center status and includes three parameters:

1. Choosing the appropriate equipment: the server utilization ratio. This ratio represents the percentage of the computer's resources that have been used. Using less than the full power of computing equipment is inefficient, with most organizations using less than 10–20 % of their resources. Virtualization and consolidation of computing equipment can increase computing equipment efficiency by up to 90 %.

2. Best practices: the data storage system utilization rate. This metric measures the amount of unused disk space; unused storage capacity is also linked to inefficient spending, while low storage utilization is linked to excess capacity allocated during deployment or poor storage resource management. For disk arrays, 40 % is currently considered a good figure. However, by implementing methods like as-needed provisioning, companies can boost efficiency by as much as 80 %. Unfortunately, many organizations may be forced to overcommit their storage capacity on purpose as they expect a rise in the amount of data soon. However, this prediction may not come true. Therefore, technology alone will not be sufficient to resolve this issue.

3. Cooling system efficiency. Air cooling systems consume approximately 50 % of the energy consumed in a typical data center. The figure may be higher if the equipment is not properly installed or runs in a suboptimal mode. A better way to remove heat from servers and use less energy for cooling is to install cabinets in the hot aisle and cold aisle configurations. The use of more efficient air conditioners and chillers that use advanced technology, such as variable blade speed pumps and fans, air and water economizers, and so on, is encouraged in data center implementation. If the efficiency of the equipment is watched, annual cooling costs can be cut by 25 %, which saves electricity.

System analysis serves as the methodology for creating a data center model here. Let us consider the system model at three different levels of the hierarchical structure: the structural modules of the data center, their purpose and functions, and how the elements interact both within the unified system and with the external environment:

$$\mathrm{SM} = f\,(K, F, G, G_{ext}),$$

where K are the data center elements;
F is the multitude of functions that the data center performs;
G is the multitude of data center characteristics;
G_{ext} are the external environment parameters.

The characteristics of the functional tasks comprising the data center's workload have been analyzed.

The following attributes are used to classify tasks Z_t based on the nature of data processing:

- computational complexity (attribute P);
- amount of processed data (attribute U);
- response time requirements (attribute T);
- request rate (attribute X).

$$Z_t = f(P_t, U_t, T_t, X).$$

With the above four characteristics, each task can be put into a different group. This makes it possible to analyze the expected load on the data center in a way that takes into account the unique features of each task.

Typically, each group of tasks $Z_i\{F_i, U_i, T_i, X\}$ can have varying degrees of influence on the choice of data center architectural solutions. Tasks with critical parameters (high computational complexity, large amounts of data, fast response time requirements, etc.) have the most influence on design decisions. Functional tasks can therefore be ranked in order of importance using the vector of their attributes. When it comes to design decisions, the needs of all the different task groups have to be met consistently.

Scalability, reliability, security, and manageability are the four key principles that guide the details of data center architecture. The structural-functional analysis of the data center enables the identification of specific functional modules (multitudes) that constitute the fundamental set for building the data center system model's structure. There are five types of multitudes:

1. $M1$ multitude. The server complex (SC) includes the following groups of servers:
   - information resources (resource servers) store and send data to application servers (DBMS servers and file servers);
   - application servers process data according to the system's business logic;
   - information presentation servers serve as a bridge between users and application servers, such as terminal servers and Web servers;
   - service servers operate other data center subsystems, such as backup system management servers.
2. $M2$ multitude. Data warehouses (DWs) are designed to organize reliable storage of information resources and provide access to them to servers.
3. $M3$ multitude. Data communication networks (DCNs) are a part of a company's data transmission network. They use equipment like routing and switching, the organization of optical channels, user connections, and information security facilities.
4. $M4$ multitude. The infrastructure of a data center consists of a variety of different systems, including a power supply and an uninterruptible power supply, a structured cable network, air conditioning and ventilation, security and fire alarm, gas fire extinguishing and smoke removal, and room access control.
5. $M5$ multitude. Data center management system. The system was designed with the integration into the company's centralized management in mind. It will provide centralized management of all data center components as well as real-time monitoring of their status, remote equipment reconfiguration, remote software loading, end-to-end monitoring and testing of data center components, equipment status reporting, resource utilization and loading, and system performance statistics collection.

The data center is structured from ready functional modules, depending on the size of the corporate network and the center itself, the load, and the features of the requests. When there is a

rise in the amount of traffic coming through the data center, the performance of the facility can be improved by increasing the size of the server complex by adding one or more servers. Moreover, all storage systems are removable and interchangeable, and additional types of storage systems can be added, whether they be more advanced or larger, making them suitable for tasks like local backup, mirroring, and others that require a lot of memory.

The data center (*DC*) system model can therefore be formulated as the following tuple of elements, characteristics, and relationships:

$$DC = \left\langle M1\left\{SC_{\overline{i=1,A};\,\overline{j=1,B}}\right\}, M2\left\{DW_{\overline{i=1,N};\,\overline{j=1,M}}\right\}, M3, M4, M5, G\left\{\Lambda, \overline{t_{ans}}\right\}\right\rangle.$$

The data center energy consumption (P_{all}) is the total energy consumption for various needs. This can be written as:

$$P_{\text{all}} = P_{\text{IT}} + P_{\text{cool}} + P_{\text{trans}} + P_{\text{other}},$$

where P_{IT} is the energy consumption for the IT equipment,
P_{cool} is the energy consumption for the cooling system,
P_{trans} is the energy lost during transmission and transformation,
P_{other} is the energy consumption for other needs (lighting, security, etc.).

Distributed data center consumption is as follows:

$$P_{\text{all}} = P_{\text{IT}} + P_{\text{trans}} + P_{\text{cool}}\,(f\,(P_{\text{seas}}, P_{\text{TCS}})) + P_{\text{other}},$$
$$P_{\text{IT}} = P_{\text{task}} + P_{\text{CPU}} + P_{\text{mem}} + P_{\text{disk}} + P_{\text{IO}} + P_{\text{actracks}} + P_{\text{backupracks}} + P_{\text{day}},$$

where
P_{CPU} is the energy consumption for the central processor to perform tasks;
P_{mem} is the energy consumption for RAM systems;
P_{disk} is the energy consumption for disk systems;
P_{IO} is the energy consumption for the input/output system;
P_{actracks} is the energy consumption for the active racks;
$P_{\text{backupracks}}$ is the energy consumption for the backup racks;
P_{day} is the CPU energy consumption per day for the selected data processing mode;
P_{TCS} is the energy consumption for the selected type of cooling system (type of cooling system, TCS);
P_{seas} are the seasonal changes in data center energy consumption ($P_{\text{cool}} = f\,(P_{\text{seas}}, P_{\text{TCS}})$.

There are several ways to reduce standby and low utilization energy consumption in today's server platforms and processors. The focus of this paper does not extend to a review of their effectiveness and capabilities.

It is reasonable to assume that only the processor can use less energy and that a single server cannot be completely turned off when the economic impact needs to be calculated. The power-saving assumptions made above were compared and verified using the Hewlett Packard server energy consumption calculator [Hewlett Packard Enterprise. TCO and ROI Calculators. https://www.hpe.com/us/en/solutions/tco-calculators.html]. An 8-core processor and four 300 GB SAS disks in a server configuration were used for the computation. At full load, the energy consumption was calculated to be 123 W, and at idle, it was 49 W. Thus, the server consumption reduction percentage is 60 %. Consequently, a 36 % reduction in consumption appears fairly credible.

The key metrics for data center energy efficiency are summarized below.

ERE (Energy Reuse Effectiveness) measures the effectiveness of energy reuse, particularly thermal energy released during data center operation.

WUE (Water Usage Effectiveness) measures the effectiveness of using water resources for cooling in data centers.

DCCE (Data Center Compute Efficiency) measures the efficiency of data center computing and gives corporate data center owners a better understanding of how effectively computing

resources are utilized without any connection to the performance of the computer systems.

IUE (Infrastructure Usage Efficiency) is a relatively new metric that was announced in June 2019 by a group of Chinese experts called TGGC (The Green Gauge China originated based on The Green Grid). IUE will assist data center operators in lowering infrastructure maintenance costs and the environmental harm that such infrastructure causes. Furthermore, it combines power supply and distribution, cooling and refrigerant distribution (including chilled air), and rack capacity calculation by height *U*. This expands the PUE concept by adding new variables that show how much load a data center can handle.

CUE (Carbon Usage Effectiveness) is the carbon utilization efficiency that indicates the environmental friendliness of data center electricity consumption. This metric is based on how the energy sources in a certain country are set up. For example, gas will be the primary fuel source in one country, oil in another, nuclear power in a third, and green power in a fourth.

SEER (Seasonal Energy Efficiency Ratio) is a seasonal energy efficiency ratio that enables the performance of data center cooling equipment to be assessed while taking the installation location and ambient temperature typical for a given season into consideration.

DCeP (Data Center Energy Productivity) is the criterion that measures how much useful work is accomplished for each unit of energy used by the data center. DCeP exceeds PUE because it considers a wider range of variables (business peculiarities of the company owning the data center).

EDE (Electronics Disposal Efficiency) measures the recycling efficiency of data center IT equipment. The cycle of computing machines has become shorter, and operators have begun switching to space-saving and high-performance server and storage systems. As a result, the metric has become increasingly important.

The metrics listed above assist operators in not only improving energy efficiency, but also demonstrating achievements in other areas such as rational use of green energy, reduction of harmful environmental impact, and monitoring of how well an IT equipment is performing.

In light of this, the system model of a data center, when taken into account with a variety of efficiency metrics, can be written as follows:

$$DC = f\,(ERE, WUE, DCCE, IUE, SEER, DCeP, EDE).$$

This system model representation requires the use of appropriate adjustments for various data center operating conditions and technical characteristics, such as data processing and storage facilities, cooling systems, and cooling agent types. When attempting to find a solution to the challenge of increasing the energy efficiency of data centers, particular focus should be placed on the modes of operation of computing equipment as well as the utilization of special-purpose system and application software.

5. CONCLUSION

Applications based on artificial intelligence can solve many problems of managing the operation of data centers. AI methods are used at various stages of data center creation and operation, including the design stage, data center capacity planning, data center operation, and optimization stage. To optimize the processing of huge amounts of data, it is necessary to perform server load balancing, and here AI applications can provide effective load balancing between servers using predictive analytics, learning from the analysis of data accumulated during the operation of the data center. One of the biggest challenges in data center optimization is predicting failures and reducing downtime. Collection and analysis of large volumes of data and their processing based on machine learning models make it possible to predict possible problems, abnormal situations in the operation of equipment, optimize the maintenance schedule, etc.

When distributing electrical power in data centers, the largest costs are related to IT equipment, maintenance, and infrastructure. In the infrastructure part of data centers, cooling

systems consume the most electricity. Thus, reducing the energy consumption of cooling systems is the key to energy savings and reducing greenhouse emissions in data centers.

Using AI to cool data centers is a very promising idea. Due to frequent load fluctuations, as well as the influence of the external environment, data center cooling systems need smart management to realize intelligent configuration and cooling on demand. Such technologies are being introduced slowly, as they require the complex technical developments and serious investments, but they are absolutely necessary. According to the McKinsey Global Institute, more than 70 % of the world's companies will implement at least one AI technology in the next decade, and more than half will use a wide range of technologies [48]. Let us highlight the main advantages of such innovations.

First of all, this is the ability to adjust the performance of the cooling system in real time depending on changes in the load on the server. The AI system monitors the operating status of the data center, is able to issue recommendations on the operation of the cooling system, redistribute the power of the cooling system and turn off inefficient cooling zones. This reduces power consumption, eliminating the risk of data center overcooling and loss of cooling capacity. In addition, as the load of the data center increases, thanks to the switching of cooling modes and the ability to learn artificial intelligence, the energy saving efficiency will increase, which will significantly reduce energy waste and reduce greenhouse gas emissions.

An equally important advantage is the minimization of the human factor. The temperature regime is automatically regulated and the released personnel can be engaged in other important tasks. By taking over routine work, the AI apps help IT administrators focus on the more important and creative aspects of maintaining data center efficiency and making the right decisions in case of force majeure.

We propose a new approach to evaluating data center energy efficiency while focusing on two key features. First, the physical resources of the data center must be virtualized. These physical resources include communication channels (bandwidth management, etc.), computing servers, storage systems (full virtual machine, resource pool), and more. Graphical interfaces for managing virtualization clusters should be made available, making it easier for administrators to manage resources, delegate them to other units, and keep track of how they are scheduled and used. Second, special-purpose AI data processing techniques should be employed. Intelligent systems with the integration of AI into the data center infrastructure should be put in place to oversee the operation of the data center equipment, enabling the equipment planning and management minimization of potential risks, server load optimization, measurement of other devices' load, and identification of overall data center energy consumption.

REFERENCES

1. *Data Center Handbook*: *Plan, Design, Build, and Operations of a Smart Data Center*. Ed. Hwaiyu Geng P.E., Wiley, 2021. doi: 10.1002/9781119597537
2. Geng H. Sustainable Data Center. In: *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*. Wiley, 2021. P. 1–13. doi: 10.1002/9781119597537.ch1
3. Wang K., Zhou Q., Guo S., Luo J. Cluster Frameworks for Efficient Scheduling and Resource Allocation in Data Center Networks: A Survey. In: *IEEE Communications Surveys & Tutorials*. 2018. V. 20. № 4. P. 3560–3580. doi: 10.1109/COMST.2018.2857922
4. Crosby C., Curtis C. Hosting or Colocation Data Centers. In: *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*. Wiley, 2021. P. 65–75. doi: 10.1002/9781119597537.ch4
5. Bajic B., Rikalovic A., Suzic N., Piuri V. Industry 4.0 Implementation Challenges and Opportunities: A Managerial Perspective. In: *IEEE Systems Journal*. 2021. V. 15. № 1. P. 546–559. doi: 10.1109/JSYST.2020.3023041

6. Rikalovic A., Suzic N., Bajic B., Piuri V. Industry 4.0 Implementation Challenges and Opportunities: A Technological Perspective. In: *IEEE Systems Journal*. 2022. V. 16. № 2. P. 2797–2810. doi: 10.1109/JSYST.2021.3101673
7. 10 Hot Data Center Technologies and Trends to Watch in 2021. *CRN Media Network*. URL: https://www.crn.com/slide-shows/data-center/10-hot-data-center-technologies-and-trends-to-watch-in-2021 (accessed 19.06.2023).
8. Shi L., Shi Y., Wei X., Ding X., Wei Z. Cost Minimization Algorithms for Data Center Management. In: *IEEE Transactions on Parallel and Distributed Systems*. 2017. V. 28. № 1. P. 60–71. doi: 10.1109/TPDS.2016.2549016
9. Yuan H., Bi J., Zhang J., Zhou M. Energy Consumption and Performance Optimized Task Scheduling in Distributed Data Centers. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2022. V. 52. № 9. P. 5506–5517. doi: 10.1109/TSMC.2021.3128430
10. Allen M. And the title of the largest data center in the world and largest data center in US goes to… *Data Center*. 2018. URL: https://www.datacenters.com/news/and-the-title-of-the-largest-data-center-in-the-world-and-largest-data-center-in (accessed 19.06.2023).
11. Torell W. Network-Critical Physical Infrastructure: Optimizing Business Value. In: *INTELEC 05 – Twenty-Seventh International Telecommunications Conference*. 2005. P. 119–124. doi: 10.1109/INTLEC.2005.335205
12. Yuan X., Zhou X., Pan Y., Kosonen R., Cai H., Gao Y., Wang Y. Phase change cooling in data centers: A review. *Ener. Buildings*. 2021. V. 236. P. 110764. doi: 10.1016/j.enbuild.2021.110764
13. Ahmed K.M.U., Bollen M.H.J., Alvarez M. A Review of Data Centers Energy Consumption and Reliability Modeling. In: *IEEE*. 2021. V. 9. P. 152536–152563. doi: 10.1109/ACCESS.2021.3125092
14. Kosik B. Energy and Sustainability in Data Centers. In: *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*. Wiley, 2021. P. 27–63. doi: 10.1002/9781119597537.ch3
15. Shaikh A., Uddin M., Elmagzoub M.A., Alghamdi A. PEMC: Power Efficiency Measurement Calculator to Compute Power Efficiency and CO_2 Emissions in Cloud Data Centers. In: *IEEE*. 2020. V. 8. P. 195216–195228. doi: 10.1109/ACCESS.2020.3033791
16. Lin W., Wu W., Li K. Energy-Saving Technologies of Servers in Data Centers. In: *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*. Wiley, 2021. P. 337–348. doi: 10.1002/9781119597537.ch19
17. Geng C.-H. Design of Energy-Efficient IT Equipment. In: *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*. Wiley, 2021. P. 323–336. doi: 10.1002/9781119597537.ch18
18. Raja S.P. Green Computing: A Future Perspective and the Operational Analysis of a Data Center. In: *IEEE Transactions on Computational Social Systems*. 2022. V. 9. № 2. P. 650–656. doi: 10.1109/TCSS.2021.3093702
19. Kosik B. Data Center Benchmark Metrics. In: *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*. Wiley, 2021. P. 617–625. doi: 10.1002/9781119597537.ch32
20. Cao Z., Zhou X., Hu H., Wang Z., Wen Y. Toward a Systematic Survey for Carbon Neutral Data Centers. In: *IEEE Communications Surveys & Tutorials*. 2022. V. 24. № 2. P. 895–936. doi: 10.1109/COMST.2022.3161275
21. Masanet E., Arman S., Koomey J. Characteristics of low-carbon data centers. *Nat. Clim. Change*. 2013. V. 3. P. 627–630. doi: 10.1038/nclimate1786
22. Why Green Data Center Matters. *FS Community*. URL: https://community.fs.com/blog/why-green-data-center-matters.html (accessed 19.06.2023).

23. Curtis P.M. Data Center Cooling Efficiency, Concepts, & Technologies. In: *Maintaining Mission Critical Systems in a 24/7 Environment, IEEE*. 2020. P. 375–396. doi: 10.1002/9781119506133.ch12
24. Roach J. To cool datacenter servers, Microsoft turns to boiling liquid. *Microsoft*. URL: https://news.microsoft.com/innovation-stories/datacenter-liquid-cooling/ (accessed 19.06.2023).
25. Han Y., Lau B.L., Tang G., Chen H., Zhang X. Si Microfluid Cooler with Jet-Slot Array for Server Processor Direct Liquid Cooling. In: *IEEE Transactions on Components, Packaging and Manufacturing Technology*. 2020. V. 10. № 2. P. 255–262. doi: 10.1109/TCPMT.2019.2933864
26. Jung S.-M., Ricci B., Chung G. Lithium Ion Battery System in Data Centers. In: *2015 IEEE 15th International Conference on Environment and Electrical Engineering* (*EEEIC*). 2015. P. 968–973. doi: 10.1109/EEEIC.2015.7165294
27. Li S., Fu Q., Chen G., Li Y., Zhang J., Feng L., Liu J., Zhou C., Liang A., Zhou H., Ahuja N., Qiao Q. An Advanced Distributed Backup Power Design with Lithium Iron Phosphate Battery for Data Center Energy Efficiency. In: *2021 20th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems* (*iTherm*). 2021. P. 563–567. doi: 10.1109/ITherm51669.2021.9503303
28. Kao W. Renewable and Clean Energy for Data Centers. In: *Data Center Handbook*. Wiley, 2015. P. 559–576. doi: 10.1002/9781118937563.ch30
29. Ahammed M.T., Osman N., Das C., Hossain M.A., Hossain S., Kaium M.H. Analysis of Energy Consumption for a Hybrid Green Data Center. In: *2022 International Conference on Innovations in Science, Engineering and Technology* (*ICISET*). 2022. P. 318–323. doi: 10.1109/ICISET54810.2022.9775899
30. Lazaar N., Barakat M., Hafiane M., Sabor J., Gualous H. Modeling and control of a hydrogen-based green data center. *Electr. Pow. Syst. Res*. 2021. V. 199. P. 107374. doi: 10.1016/j.epsr.2021.107374
31. Liu L., Zhang Q., Zhai Z., Yue C., Ma X. State-of-the-art on thermal energy storage technologies in data center. *Ener. Buildings*. 2020. V. 226. P. 110345. doi: 10.1016/j.enbuild.2020.110345
32. Guo P., Wang S., Lei Y., Li J. Numerical simulation of solar chimney-based direct airside free cooling system for green data centers. *J. Buil. Eng*. 2020. V. 32. P. 101793. doi: 10.1016/j.jobe.2020.101793
33. Global artificial intelligence market size 2021 rise at 35.6% CAGR will grow to USD 299.64 billion by 2026. *Facts & Factors*. URL: https://www.globenewswire.com/en/news-release/2021... (accessed 19.06.2023).
34. Ahdoot A.A. Three ways artificial intelligence will revolutionize data centers. *Data Center Knowledge.* URL: https://www.datacenterknowledge.com/industry-perspectives/... (accessed 19.06.2023).
35. Evans R., Gao J. DeepMind AI reduces google data centre cooling bill by 40%. *Google DeepMind*. URL: https://www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40 (accessed 19.06.2023).
36. Kliazovich D., Pecero J.E., Tchernykh A., Bouvry P., Khan S.U., Zomaya A.Y. CA-DAG: Modeling communication-aware applications for scheduling in cloud computing. *J. Grid Comput*. 2016. V. 14. P. 23–39. doi: 10.1007/s10723-015-9337-8
37. Ahmad R.W., Gani A., Hamid S.H., Shiraz M., Yousafzai A., Xia F. A survey on virtual machine migration and server consolidation frameworks for cloud data centers. *J. Netw. Comput. Appl*. 2015. V. 52. P. 11–25. doi: 10.1016/j.jnca.2015.02.002
38. Armenta-Cano F., Tchernykh A., Cortés-Mendoza J.M., Yahyapour R., Drozdov A., Bouvry P., Kliazovich D., Avetisyan A., Nesmachnow S. Min_c: Heterogeneous

concentration policy for energy-aware scheduling of jobs with resource contention. *Program. Comput. Soft*. 2017. V. 43. P. 204–215. doi: 10.1134/S0361768817030021

39. Muraña J., Nesmachnow S., Armenta F., Tchernykh A. Characterization, modeling and scheduling of power consumption of scientific computing applications in multicores. *Cluster Comput*. 2019. V. 22. № 3. P. 839–859. doi: 10.1007/s10586-018-2882-8
40. Feng H., Deng Y., Li J. A global-energy-aware virtual machine placement strategy for cloud data centers. *J. Syst. Architect*. 2021. V. 116. P. 102048. doi: 10.1016/j.sysarc.2021.102048
41. Helali L., Omri M.N. A survey of data center consolidation in cloud computing systems. *Comput. Sc. Rev*. 2021. V. 39. P. 100366. doi: 10.1016/j.cosrev.2021.100366
42. Khoshkholghi M.A., Derahman M.N., Abdullah A., Subramaniam S., Othman M. Energy-Efficient Algorithms for Dynamic Virtual Machine Consolidation in Cloud Data Centers. In: *IEEE*. 2017. V. 5. P. 10709–10722. doi: 10.1109/ACCESS.2017.2711043
43. Uddin M., Darabidarabkhani Y., Shah A., Memon J. Evaluating power efficient algorithms for efficiency and carbon emissions in cloud data centers: A review. *Renew. Sust. Energ. Rev*. 2015. V. 51. P. 1553–1563. doi: 10.1016/j.rser.2015.07.061
44. Liang B., Wu D., Wu P., Su Y. An energy-aware resource deployment algorithm for cloud data centers based on dynamic hybrid machine learning. *Knowl.-Based Syst*. 2021. V. 222. P. 107020. doi: 10.1016/j.knosys.2021.107020
45. Wodecki B. Alibaba opens AI-focused data center in Germany. *Data Center Knowledge*. 2022. URL: https://www.datacenterknowledge.com/cloud/alibaba-opens-ai-focused-data-center-germany (accessed 19.06.2023).
46. Burgess M. Google's DeepMind trains AI to cut its energy bills by 40%. *Wired*. 2016. URL: https://www.wired.co.uk/article/google-deepmind-data-centres-efficiency (accessed 19.06.2023).
47. Patterson D. Good news about the carbon footprint of machine learning training. *Google AI Blog*. 2022. URL: https://ai.googleblog.com/2022/02/good-news-about-carbon-footprint-of.html (accessed 19.06.2023).
48. McKinsey report: Two AI trends top 2022 outlook. *Venture Beat*. 2022. URL: https://venturebeat.com/ai/mckinsey-report-two-ai-trends-top-2022-outlook/ (accessed 19.06.2023).

Received 29.02.2023.
Revised 19.06.2023.
Published 23.06.2023.