Virtual Reality & Intelligent Hardware

• Review •

August 2022 Vol. 4 No. 4: 279–291
10.1016/j.vrih.2022.05.003

Digital twin intelligent system for industrial internet of things-based big data management and analysis in cloud environments

Christos L. STERGIOU, Kostas E. PSANNIS*

Department of Applied Informatics, University of Macedonia, Greece

Received 3 April 2022; **Revised** 22 April 2022; **Accepted** 24 May 2022

Abstract: This work surveys and illustrates multiple open challenges in the field of industrial Internet of Things (IoT)-based big data management and analysis in cloud environments. Challenges arising from the fields of machine learning in cloud infrastructures, artificial intelligence techniques for big data analytics in cloud environments, and federated learning cloud systems are elucidated. Additionally, reinforcement learning, which is a novel technique that allows large cloud-based data centers, to allocate more energy-efficient resources is examined. Moreover, we propose an architecture that attempts to combine the features offered by several cloud providers to achieve an energy-efficient industrial IoT-based big data management framework (EEIBDM) established outside of every user in the cloud. IoT data can be integrated with techniques such as reinforcement and federated learning to achieve a digital twin scenario for the virtual representation of industrial IoT-based big data of machines and room temperatures. Furthermore, we propose an algorithm for determining the energy consumption of the infrastructure by evaluating the EEIBDM framework. Finally, future directions for the expansion of this research are discussed.

Keywords: Machine learning; IoT; Big data; Cloud computing; Management; Analytics; Digital twin Scenario; Energy efficiency

Citation: Christos L. STERGIOU, Kostas E. PSANNIS. Digital twin intelligent system for industrial internet of things-based big data management and analysis in cloud environments. Virtual Reality & Intelligent Hardware, 2022, 4(4): 279–291

1 Introduction

This study surveys and integrates emerging and novel technologies aimed at providing a more energy-efficient environment for managing and analyzing industrial Internet of Things (IoT)-based big data in a cloud environment. Specifically, by developing an intelligent virtual system framework based on the concepts of reinforcement and federated learning, we can achieve better and faster management and analysis of big data (BD) in the cloud environment with low power consumption. The proposed system is simulated as a virtual representation scenario along with the industrial IoT-based big data generated using CloudSim software.

BD is a complex concept involving many factors, such as volume, velocity, variety, veracity, and value. Each value corresponds to one of the five Vs of big data[1,2]. Furthermore, BD often requires predictive analytics or

***Corresponding author,** kpsannis@uom.gr

certain advanced approaches to extract value from data[3].

Cloud computing (CC) offers an environment for accessing information and data from any place at any time, with no limitations on the hardware requirements of the user. Therefore, cloud providers must have large and efficient data centers to manage the high demand for resources from users. Thus, there is a need for more energy-efficient data centers that simultaneously offer high computing power. Furthermore, CC employs three specific data models for its functions: SaaS, PaaS, and IaaS. These models offer opportunities to both cloud providers and users[4–6].

The management of industrial IoT-based big data is an emerging field of study with numerous challenges, and research on achieving more efficient and productive management of the data is being actively conducted[7–9]. The IoT data used in this study is a virtual representation of machines and room temperatures. BD analytics is based on techniques that investigate the field of industrial IoT-based BD management to produce more useful information from the data[10]. Large databases that store multiple types of data need to be analyzed to obtain information aimed at making everyday life easier. These large database centers are usually located alongside cloud data centers, and thus, more energy is consumed. In addition, to achieve easier data management and a more energy-efficient environment, novel learning techniques such as reinforcement and federated learning can be applied.

Federated learning (FL) is an innovative technique that provides new avenues for the operation and management of data in collaborative environments. The FL scenario algorithms consider communication with edge devices taking place over the networks, which in most cases are unreliable, as well as the limitations on upload speeds[11,12]. Consequently, FL can significantly reduce security and privacy risks by restricting the attack to only the device, thereby safeguarding the device's cloud[13,14]. Furthermore, in industrial IoT-based BD, which is the primary technology apart from the CC on which this study focuses, the number of training examples is typically too large to be stored on one computer, and therefore, we need to divide the computation among many computers. Moreover, an effective method to run and manage systems based on the idea of FL and machine learning systems in general is to use a virtual representation of a system to obtain a useful solution[15]. Such virtual systems are based on the idea of digital twin scenarios through which we can operate a cloud system as a virtual representation for managing and analyzing industrial IoT-based BD produced through industrial procedures. In this study, to achieve the digital twin scenario, we used the CloudSim software.

In conclusion, the key contributions of this study can be summarized as follows:

1. Open challenges and issues in the field of energy-efficient industrial IoT-based big data management and analysis in cloud environments are surveyed.

2. An architecture that attempts to combine the features offered by several cloud providers to achieve an energy-efficient industrial IoT-based big data management framework (EEIBDM) established outside of every user in the cloud is proposed.

3. Perspectives on the problem of resource allocation in interactions with the cloud environment to achieve an optimal decision are formulated.

4. An algorithm that determines the energy consumption of the CPU through the evaluation of the EEIBDM framework is proposed.

The remainder of this paper is organized as follows. Section 2 details the relevant research in the field of energy-efficient BD management in cloud environments, the artificial intelligence scenarios of cloud infrastructures, and the artificial intelligence techniques of BD analytics in cloud environments. Section 3 provides details of the proposed approach and architecture of the proposed system. Section 4 discusses the evaluation of the proposed system and presents the proposed algorithm, modified CloudSim architecture, experimental results, and topology of our scenario. Finally, Section 5 concludes the study and outlines future directions.

2 Review of related work

In this study, several previous studies were reviewed and analyzed in terms of related topics. The following subsections illustrate the work that has contributed significantly to our study.

2.1 Efficient big data management in the cloud

In recent years, several works have been published in the field of efficient BD Management in cloud environments. Thus, for the purpose of our research, we have studied and analyzed previous research that was done in the field of efficient BD management in cloud environment[16–18]. The following paragraphs present previous research related to our study.

Aujla and Kumar presented MEnSuS, which is an efficient system for the sustainability of cloud data centers in an edge-cloud environment, using SDN for energy management[16]. The proposed scheme is a workload-based approach to workload classification. Al-Dulaimy et al. investigated how to achieve new strategies for energy-efficient cloud data centers by creating new designs and implementing virtual machine (VM) management and proposed a novel model for solving VM placement problems[17]. Khan et al. proposed a wide range of heuristic and meta-heuristic VMC algorithms because it is an NP-hard problem that aims to achieve near-optimality[18].

2.2 AI cloud scenarios

Furthermore, regarding the novel scenarios of artificial intelligence cloud systems and the use of data analytics, several remarkable works are related[19–21]. The following paragraphs present research papers related to this study.

Weber et al. presented an approach to rollback cloud management, considering the particular needs of users of CC resources wishing to manage the resources[19]. Their proposed approach wrapped the cloud management API and used A.I. planning techniques to find a proper undo sequence. Brown and Kauchak discussed and shared novel educational approximations that teach or leverage artificial intelligence and its many subsections, covering computer vision, robotics, natural language processing, machine learning, and others in all layers of education[20]. Rad et al. presented the Cloud-eLab platform, which is an open and interactive cloud-based learning platform for AI. Thinking, intends to infuse two aspects: i) deep and broad learning, and ii) cognitive and adaptation learning notions for education[21].

2.3 AI big data analytics in the cloud

Moreover, several remarkable works are associated specifically with artificial intelligence systems of big data analytics in cloud environments[22–26]. The following paragraphs are related to our research.

Wu et al. reviewed the historical perspectives of the term “big data”. Through their review and the correlated analytics, they conclude that BD does not consist of only 3Vs, but it could be divided into 32 Vs[22]. Specifically, Wu et al. concluded that nine Vs cover the cardinal motivation inside the BD term, which is to merge business intelligence (BI) count on various hypotheses or statistical models so that BD Analytics (BDA) can assist decision-makers in making useful predictions for fatal decisions or researching results. Ahmed et al. investigated recent advances in BDA for IoT systems, such as the key requirements for managing BD and for activating analytics in an IoT environment[23]. Lee et al.[24] considered the 5C architecture proposed by a previous work of Lee et al.[27] and proposed an insight into the ongoing AI technologies and the ecosystem needed to harness the power of AI in industrial applications. Wan et al. presented a vertically integrated four-level CaSF (cloud-assisted smart factory) architecture[25]. With this proposal, Wan et al. aimed to highlight the role and contingency of CC and AI in ameliorating the smart factories’ performances, such as system efficiency, flexibility, and intelligence, and they completely explained and summarized the AI application in a cloud-assisted smart factory (CaSF). Khan et al. explored the current research, challenges, open problems, and future research directions for several problems that need to be confronted and risks that need to be moderated before practical applications of this synergistic model can be popularly used[26].

2.4 Comparative analysis

Regarding the research conducted in previous works, we analyze our findings in Table 1. Specifically, Table 1 illustrates the basic model features of previous works in the related field, in comparison with our proposal. In particular, most of the studies do not contribute to technologies such as FL and reinforcement learning. In addition, most of them are not associated with IoT applications or the data they produce. Another conclusion that we can observe is that most of the related studies implement platforms. Few of them try to find solutions for the open issues and needs of cloud users, and more energy-efficient systems for the operations. On the other hand, most related works have contributed to the broader sense of artificial intelligence regarding novel techniques, applications, interfaces, and platforms. Consequently, regarding the findings listed in Table 1, we can consider that our proposed scenario tries to collaborate with novel scenarios such as federated and reinforcement learning systems, and we embed techniques that implement a more energy-efficient system that aims to offer more useful and efficient management of industrial IoT big data, combining CC, BD, and IoT in a novel framework. Table 2 illustrates the definitions of the abbreviations in Table 1.

Table 1 Comparison of challenges, issues, and proposals of existing studies

Work	NC	BPT	EES	AIT	AIA	AII	AIC	RM	VC	RC	IS	IA	BA	BM	FM	RM	P	A	F
Aujla & Kumar[16]	M	L	H	L	L	L	L	M	–	●	L	M	L	H	–	–	●	–	–
Al-Dulaimy et al.[17]	L	L	H	L	M	L	M	H	●	●	M	L	L	H	–	–	●	●	–
Khan et al. [18]	M	H	M	L	L	L	H	H	●	–	M	L	M	M	–	●	–	–	●
Weber et al. [19]	L	H	L	M	H	L	H	L	●	●	M	L	L	H	●	–	●	–	–
Brown & Kauchak [20]	M	M	L	H	H	H	H	M	–	●	M	M	H	H	●	–	●	–	●
Rad et al. [21]	M	M	L	H	H	M	H	M	●	–	H	L	L	L	●	–	●	–	–
Wu et al. [22]	L	H	L	H	H	H	M	L	–	●	L	L	H	H	–	●	–	–	●
Ahmed et al. [23]	L	M	L	M	M	L	L	M	–	●	M	H	H	H	–	–	●	–	–
Lee et al. [24]	M	M	M	H	H	M	H	M	●	–	M	L	L	H	–	●	●	●	–
Wan et al. [25]	M	M	H	M	H	M	H	H	●	–	H	L	M	M	–	–	–	●	–
Khan et al. [26]	L	M	M	H	H	L	M	M	●	–	M	L	M	M	–	–	–	●	–
Proposed Model	M	H	H	H	M	M	H	H	–	●	H	M	M	H	●	●	–	●	●

Table 2 Parameter definitions

NC: Needs of Cloud Computer users	**RR:** Resources management	**FM:** Federated methods
BPT: Big Data processing techniques	**VC:** Virtual Cloud	**RM:** Reinforcement methods
EES: Energy-efficient scenario	**RC:** Real Cloud	**P:** Propose platform
AIT: AI techniques	**IS:** Integration Scenarios	**A:** Propose architecture
AIA: AI applications	**IA:** IoT applications	**F:** Propose a framework
AII: AI interfaces	**BA:** Big Data analytics	**H:** High **L:** Low
AIC: AI Cloud-based platform	**BM:** Big Data management	**M:** Medium

3 Approach of proposed system

The evaluation of our proposed work is presented in this section.

3.1 Energy efficiency in data center

It is a fact that the large, industrial data center infrastructure assumes a huge energy consumption cost for the resources of the infrastructure, which leads to a considerable increase in environmental costs. This is a major issue regarding the carbon footprint and the energy cost of cloud systems. To reduce energy consumption, intelligent mechanisms must be built with the ability to be managed across different heterogeneous machines.

In our previous study on literature review, to achieve energy efficiency, we had to integrate mechanisms of reinforcement learning and federated learning to achieve a cloud system capable of decreasing the consumption of infrastructures not in use.

The most widely accepted unit of measurement for energy efficiency in a data center is power usage effectiveness (PUE). The PUE has already been defined and introduced by Armbrust et al.[28] as a green grid component, as well as the ratio of the total power used in a data center installation to the power supplied to the IT equipment. Specifically, the *PUE* can be expressed using the following equation.

$$PUE = \frac{TFP}{ITEP} \tag{1}$$

Here, the value of *TFP* represents the total facility power, which denotes the data centers' entire power delivered. In contrast, *ITEP* represents the IT equipment power, which illustrates the energy facilities consumed by the equipment used to manage, transfer, process, operate, store, and route data. Because the experiment was previously analyzed by Koutitas and Demestichas[29], the result of Equation 1 mostly emphasizes the energy consumption of the cloud data center's IT equipment, which accounts for 30% of the entire data center. Thus, the *PUE* of Equation 1 can be formulated as follows.

$$PUE = \frac{NIT_{pc} + (CPU_{pc} + NonCPU_{pc})}{(CPU_{pc} + NonCPU_{pc})} \tag{2}$$

Here, to produce the value of *PUE*, we calculate the sum of the value of NIT_{pc}, which represents the *NonIT* equipment power consumption (30% of the total), and the summary of CPU_{pc}, which represents CPU power consumption (40% of the total), and $NonCPU_{pc}$, which represents *NonCPU* power consumption (10% of the total), divided by the sum of CPU_{pc} and $NonCPU_{pc}$. The energy consumption of the CPU can be obtained from our proposed method presented in Algorithm 1 and represents the use of the resource allocation scenario provided by our energy-efficient industrial IoT-based big data management (EEIBDM) framework. Finally, the overall-high value of CPU power consumption in the data center produces a high expenditure on the cooling system; however, it is not necessary in our scenario.

3.2 Reinforcement learning contribution to energy-efficient resource allocation

Reinforcement learning (RL) can be defined as "*a learning process through interactions with a dynamic environment, which produces the optimal control policy for a given set of situations without requiring knowledge of the environment*"[30,31]. Several functions endlessly extract the reward through the learning process of the system. Furthermore, the RL count has two necessary functions:1) trial-and-error retrieval, and 2) delayed reward.

Thus, the formulation of the problem of energy-efficient resource allocation in a cloud environment, which aims to be addressed by RL, can be introduced as an optimization problem concerning the already well-known *Markov Decision Processes*. Our research goal lies in the formulation of the angles of the resource allocation issue while interacting with the cloud environment to attain an optimal decision.

3.3 Federated cloud system modeling

We attempted to model a cloud coordinator entity as a requirement for federating multiple clouds. The cloud coordinator (ClCo) is responsible for managing and monitoring the inner state of a cloud datacenter entity, except for communication with end-users and other datacenters. The information received from the ClCo as a part of the monitoring process, which is ongoing during the pilot period, is used to make decisions related to inter-cloud provision. ClCo functionality might be defined as similar to the functionality offered by large businesses. As a result, when an engineer of a cloud system demands federation of services from several cloud providers, the development of a ClCo will be required. Thus, aspects associated with communication and negotiations with former entities are isolated from the core of the data center to enable an entity to manage the federation scenario of cloud data centers. Consequently, the CloudSim operation provides every cloud developer with the ability to speed up the use of application services by performing tests through an entity such as ClCo.

4 Evaluation of proposed approach

Regarding the study, aiming to succeed in creating a novel efficient system for industrial IoT-based BD management and analysis in a cloud environment, we used RL and FL techniques to formulate and design the architecture of the proposed system.

Industrial IoT-based data management and analysis in cloud infrastructure has become popular in recent years because of the help of software ‘infrastructure, which efficiently supports the operation of datacenters and cloud datacenters. Owing to the maximum need for hardware infrastructure and the need to continuously update the data centers, providers intend to host all their infrastructure in data centers that can support as many customers as they can, thus adopting virtualization. Virtualization is a new technique in which users are given virtual platforms, rather than physical ones, aiming to unravel many operational and maintenance issues in a data center. Virtualization can be characterized as an effective way to offer management solutions for dynamic resources in a cloud environment.

Virtual machines (VMs) can be used through virtualization. A VM is an identical, isolated execution environment on a single computer, which emulates the host computer. As a result, this gives the user the illusion of having a physical machine. These VMs can be used in emulator simulations. An emulator is used to simulate a hardware platform, generally to allow running multiple operating systems simultaneously and to support foreign code on a given platform. The emulator used in our research was CloudSim, which operates on Eclipse.

The proposed architecture scenario attempts to combine the features offered by several cloud providers to achieve a sustainable and energy-efficient industrial IoT-based big data management framework (EEIBDM) established outside of every user in the cloud. The proposed system architecture relies on the IaaS and PaaS models of the cloud provided by the cooperative cloud providers to meet metrics such as CPU resources, memory amount, storage availability, and system performance in terms of execution time. Inferring from Figure 1, each different type of user could use the cloud infrastructures of the various cooperative cloud providers through the EEIBDM, which offers three major advantages:1) energy-efficient resource allocation, 2) data center manager/analyzer, and 3) cloud infrastructure resource monitor. Each type of user can access the system framework via PC, laptops, mobile devices, etc., and has the same type of access regarding the permissions granted to access the cloud infrastructure. The proposed system acts as an intermediary to provide better management and secure access to each user when trying to use the cloud.

4.1 Reinforcement based cloud evaluation resource allocation

Based on the study, we know that a service level agreement (SLA) administrator investigates the CPU usage associated with all servers running in the abbot data center for the distribution of VM cloud environments to ensure the SLA metric distribution condition. The calculation can be expressed as follows.

$$EC_n = \int_{t0}^{t1} F(u(t))\mathrm{d}t \quad n = 1, 2, ..., k \tag{3}$$

Here, EC_n represents the overall energy consumption of a specific host that functions in the overall cal-

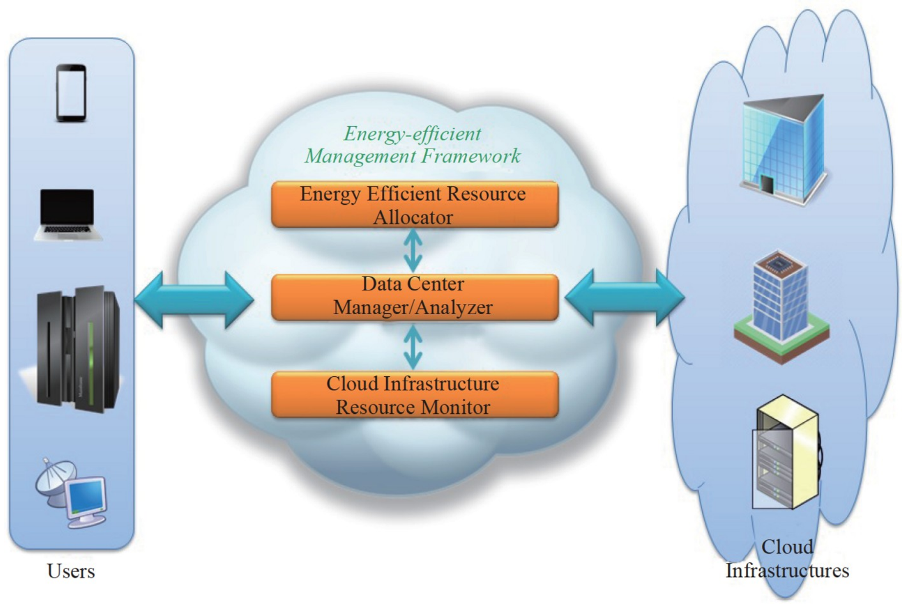

Figure 1 System architecture.

culation time. Moreover, CPU utilization is denoted by $u(t)$, the period of the overall calculation time for each host is denoted by n, and its range is defined by the total number of hosts contributing to the datacenter, starting from 1 to k.

$$EC = \sum_{n=1}^{k} EC_n \tag{4}$$

Resulting in Equation 3, the calculation of the total quantity of energy consumption of the data center representing the energy consumption of all the contributed hosts in the data center, which can be better expressed using Equation 4.

As is evident from Table 3, the hardware setup of each cloud datacenter is structured.

Table 3 CloudSim system configuration-reinforcement cloud-based evaluation

Data Center (Host)	Virtual Machine
12GB RAM memory	512MB RAM memory
2TB storage memory	20GB storage memory
2×CPU with 1000 MPIS capacity	1×CPU with 1000 MPIS capacity
Time-shared VM scheduler	Time-shared Cloudlet scheduler

4.2 Evaluation of resource allocation in EEIBDM

Algorithm 1 was introduced as a novel resource-allocation algorithm. This algorithm was embedded and tested using the CloudSim toolkit software. All aspects of the proposed framework are included as part of an extensive heuristic in the CloudSim toolkit. As already mentioned in Section 3, CloudSim, as virtualization software, could consist of a scalable simulation framework that allows innovative support for modeling, simulation, and experimentation of virtualized data centers in cloud environments, as well as cloud management services for all components such as VMs, memory, storage, and bandwidth, under various capabilities, configurations, and domains. Finally, CloudSim can support characteristics that model and simulate environments based on large-scale clouds, resource allocation policies of energy-efficient scenarios, service brokers, virtualization techniques, federated cloud systems of CPs, and established network connections.

The computational complexity of the proposed algorithm relies on the multiple scenarios that we ran on CloudSim. Each scenario differs in time duration and the amount of data used. As observed from the experimental results in the next section, our proposal was tested for five days and for multiple types of VMs used and produced data.

Considering the characteristics provided by our proposed method, CloudSim can exploit novel construct heuristics to assess the performance obstacles associated with service delivery and provisioning policies in

Algorithm 1 EEIBDM's framework resource allocation scenario

```
The inputs of the method are:
    √  mber of the different hosts operates in the data center initialized: NoHost
    √  The number of the various VMs operates in the data center initialized: NoVM
    √  The CPU's workload value counts on the different users of the system per second: cpuw
    √  The discount factor of the system: dfs
    √  The particular upper limit of the learning process: U
The method exports as output:
    √  optimized distribution of the used VMs: overall allocation
Proposed Method:
initialize Host(NoHost) // create Hosts operating in data center with specific features
initialize VM(NoVM) // create VMs operating in data center with specific features
initialize CPUW(cpuw) // initialize CPU's workload and data center components
create Environment() // set up state set S, action set A and initialize K values and F values
for VM ϵ NoVM // each VM contained in Number of VMs
   for Host ϵ NoHost // each Host contained in Number of Hosts
      Sb = {edc, h, vm} // convey values of energy of data center, host and Vm of the specific state Sb
      for i, i=0,1,2,3,…,U
         Ai = A ϵ max_A * K_i(S_i, A')
         with Ai count Ai+1
         recompense Fi+1
         K_{i+1}(S_i, A_i) ← F + [dfs · max_{A'} F · K_i(S_{i+1}, A)]
         // update the existing value
         Si = Si+1 // distribute the next host
      end
   end
   return h
   distribute(Host, VM) // allocate new host and VM
end
return overall allocation
```

resource management techniques. Therefore, the existing architecture of CloudSim software supports cloud infrastructure service management, but unfortunately does not consider the energy consumption of a data center or the *PUE* value.

Figure 2 illustrates the existing architecture of the CloudSim software integrated with our proposed EEIBDM state to achieve an energy-efficient resource allocation service through the existing architecture. As shown in Figure 2, it is placed in the middle of the core of the CloudSim setup component.

4.3 Virtual environment task

To produce a more efficient system, we modeled and embedded a federated cloud network in CloudSim. Thus, we modeled and ran a system of three CP federations and a connection to a user broker. Each CP institutes a sensor that is responsible for dynamically detecting the availability of information related to the data hosts. Subsequently, the measurements of this sensor were delivered to the ClCo, where the produced information was utilized in undertaking load-migration decisions. As a result, this system performs transmigration of the available VMs through the cooperative CPs, taking into account the possibility that the initial CP cannot provide the requested number of available VM slots. The topology of this scenario is shown in Figure 3, which demonstrates the cloud provider federation. As shown in Figure 3, a user broker can access the cloud space through a network (e.g., wireless 4G/5G network). The cloud coordinator that coordinates the access of each user and the permission of each user attempts to find a suitable cooperative CP to serve the user in a better and more reliable way.

The model components of the proposed CloudSim simulation for the aforementioned scenario are listed in Table 4. The performance results of the federated cloud simulation in CloudSim for our proposed cloud scenario are listed in Table 5.

Assuming the operation of ClCo and the previously proposed methods, the cloud computing architecture of the system can be represented as shown in Figure 4. Figure 4 shows the EEIBDM system, including techniques

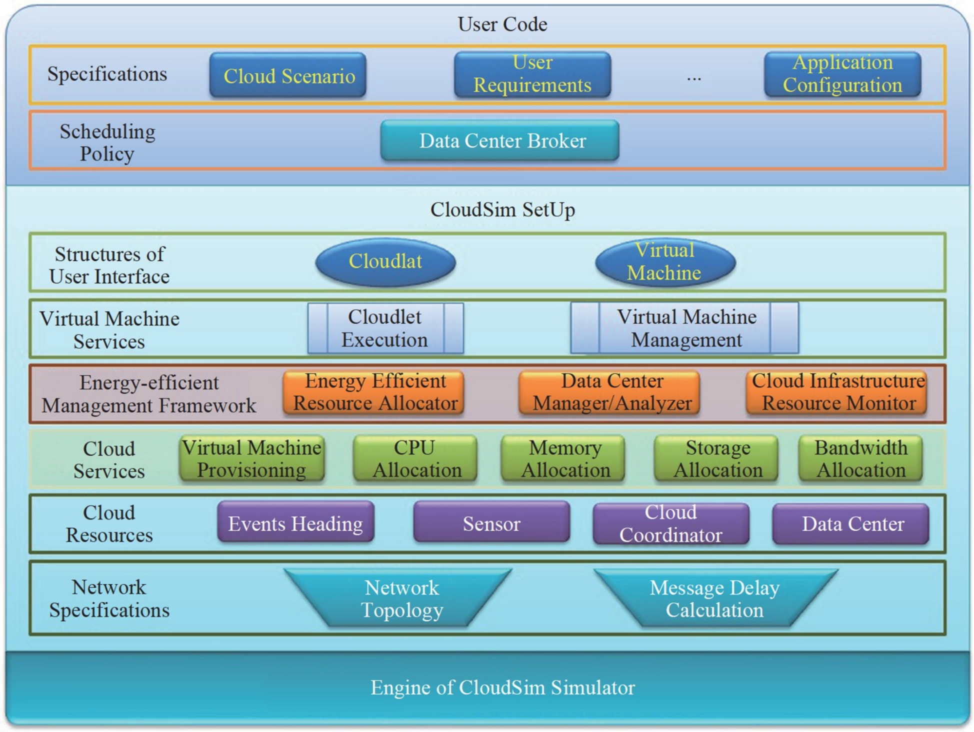

Figure 2 CloudSim architecture emerges with EEIBDM.

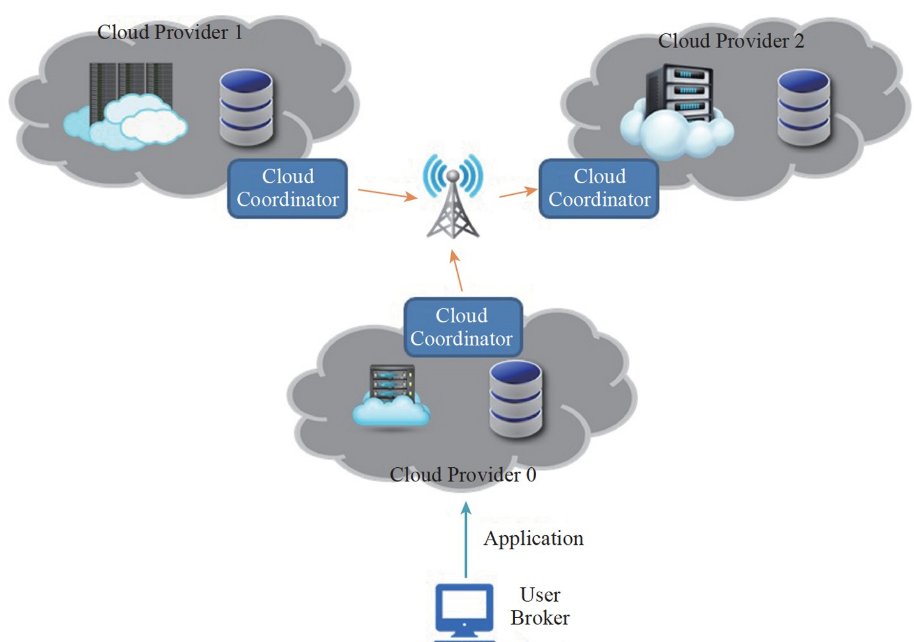

Figure 3 Federated cloud data center topology.

that rely on the data center, ClCo, and sensor components. Through the embedded sensors, the ClCo can monitor the performance of each active VM during this time. Thus, the virtual machine manager (VMM) obtains real-time data and then uses these data to perform a particular resize of the VMs needed. Finally, the ClCo allocates the VMs by applying VM migration and changes the power state of each node, following the rules of resource utilization. Each autonomous system consisted of a sensor, VMM, and other components, which could automatically serve the user's demand at the time the ClCo chooses it. The energy required is divided into several VMs; thus, the required energy is less than the original energy.

5 Experimental results of reinforcement cloud evaluation

The experimental results of the reinforcement cloud simulation in CloudSim are presented and analyzed in this section.

Figure 5 shows the effects of the energy consumption (EC) of the operation on the need for VMs for each day. More specifically, it shows how the energy of the data center is consumed by serving the requests from VMs in the 5-day schedule, for five states of the number of VMs used. As shown in Figure 5, as the number of VMs increased, the energy consumption in the data center increased. Additionally, as an overall statistic, we can state that for 50VMs, the EC is below 12kW/h, for 100VMs, it is below 15kW/h, for 150VMs, it is under 20kW/h, for 200VMs, it is under 25kW/h, and for 250VMs, it is less than 29kW/h. Moreover, we can conclude that during the past days, EC has decreased.

Table 4 CloudSim configuration–federated cloud set-up

Data Center (Host)	Virtual Machine (x50)
100 computing hosts	1×VM = 1×Cloudlet
12GB RAM memory	512MB RAM memory
2TB storage memory	20GB storage memory
2×CPU with 1000 MPIS capacity	1×CPU

Table 5 Performance results of federated cloud obtained using cloudsim

Performance metrics	Federation results	Non-Federation results
Average turnaround time (sec)	4241.45	8782.9
Makespan (sec)	7653.62	14609.41

Figure 6 shows that the value of PUE is decreasing in value as the need to use more VMs grows. Thus, the major goal of achieving energy-efficient usage when more VMs are required, is reached. According to Figure 6, when there is a need for 50 VMs the PUE is under 1.95 for all five days. Additionally, for 100VMs the PUE is below 1.89, for 150VMs, it is below 1.80, for 200VMs, it is under 1.70, and for 250VMs it is no more than 1.65. According to the literature, the range of an energy-efficient value of PUE is between 1 and 2. Consequently, the proposed algorithm and method achieve the energy efficiency level of a data center regarding the results demonstrated in Figure 6.

Figure 7 shows the percentage of SLA violations performed during five days of operation of the reinforcement cloud system in CloudSim. It was demonstrated that the higher the number of VMs, the more the percentage of SLA violations increased, so they were analogous values. An increase of approximately 5% in SLA violation resulted in 50VM allocations, while an increase of 10% in SLA violation resulted in 100VM

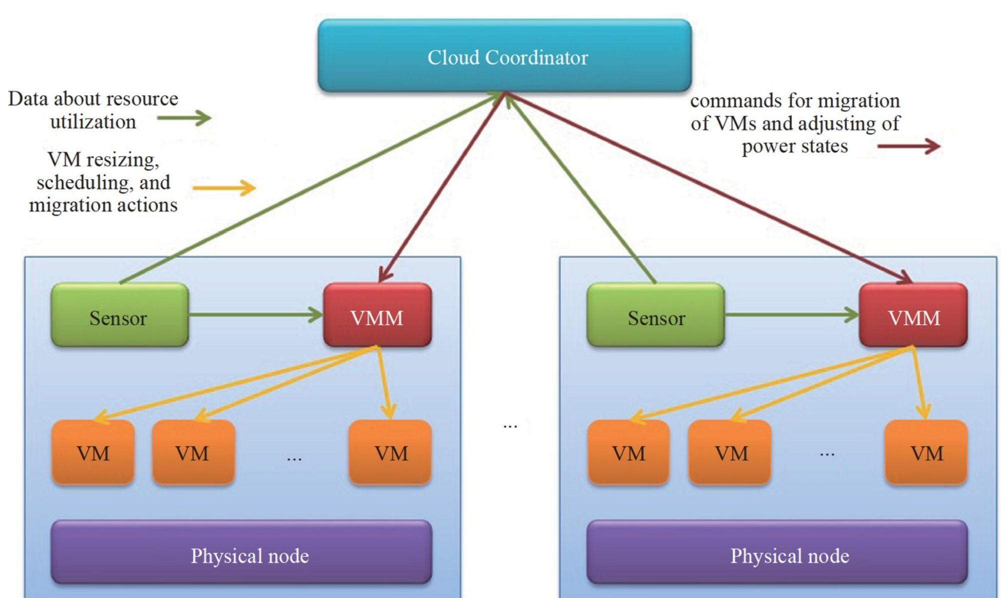

Figure 4 Energy-conscious management architecture.

allocations, 16% for 150VM allocations, 20% for 200VM allocations, and 27% for 250VM allocations.

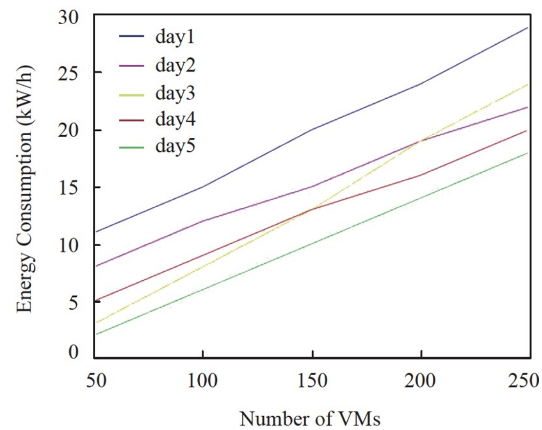

Figure 5 Energy consumption during five days of operation of the reinforcement cloud system in CloudSim.

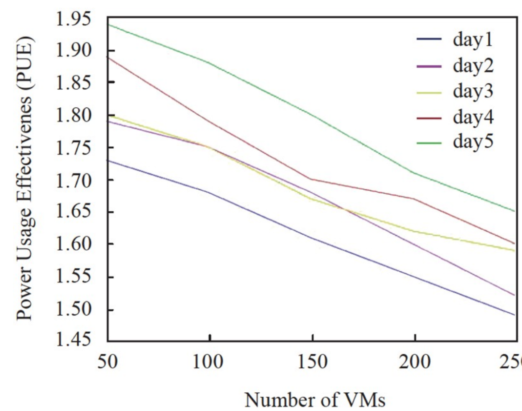

Figure 6 Power usage effectiveness during five days of operation of the reinforcement cloud system in CloudSim.

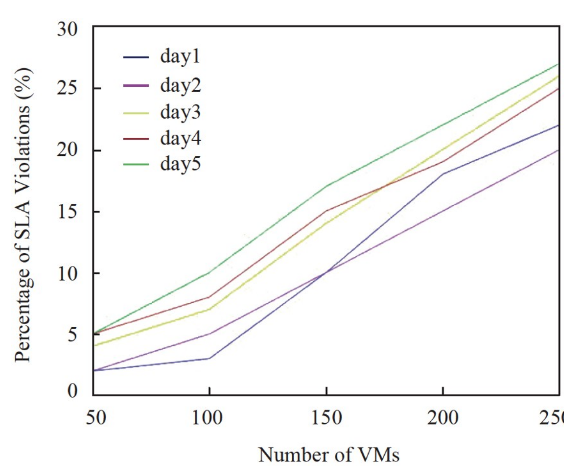

Figure 7 Percentage of SLA violation during five days of operation of the reinforcement cloud system in CloudSim.

6 Conclusion

This study surveyed multiple open challenges and problems in the field of sustainable industrial IoT-based big data management and analysis in cloud environments, particularly, the aspects and challenges arising from the fields of machine learning scenarios of cloud infrastructures, artificial intelligence techniques of industrial IoT-based BD analytics in cloud environments, and federated learning cloud systems. Considering that reinforcement learning is a novel technique that allows large data centers such as cloud data centers to influence a more energy-efficient resource allocation, we propose an architecture that attempts to combine the features offered by several cloud providers to emerge and achieve an energy-efficient industrial IoT-based big data management framework (EEIBDM) established outside of every user in the cloud environment. As a result, the major goal of this study is the formulation of various aspects of the resource allocation issue, considered from the reinforcement learning scenario, while interacting with the cloud environment to achieve an optimal decision. To achieve this, we propose an algorithm for delivering the energy consumption of the CPU through the evaluation of the EEIBDM framework.

As a case study for the future, we plan to incorporate security and privacy aspects into our proposed system framework to achieve an energy-efficient and secure cloud-based management and analysis environment based on industrial IoT, with the help of innovative techniques of reinforcement and federated learning. Thus, this proposed framework could be used in places such as hospitals, schools, and repositories of legal cases to have a more secure environment, in addition to the most energy-efficient environment. These are future directions that extend our proposal and plan to be investigated in future research.

Declaration of competing interest

We declare that we have no conflict of interest.

References

1 Psannis K E, Stergiou C, Gupta B B. Advanced media-based smart big data on intelligent cloud systems. IEEE Transactions on Sustainable Computing, 2019, 4(1): 77–87
DOI: 10.1109/tsusc.2018.2817043

2 Stergiou C, Psannis K E. Efficient and secure BIG data delivery in Cloud Computing. Multimedia Tools and Applications, 2017, 76(21): 22803–22822
DOI: 10.1007/s11042-017-4590-4

3 Jiang X, Ge Z. Information fingerprint for secure industrial big data analytics. IEEE Transactions on Industrial Informatics, 2022, 18(4): 2641–2650
DOI: 10.1109/tii.2021.3104056

4 Stergiou C, Psannis K E, Kim B G, Gupta B. Secure integration of IoT and cloud computing. Future Generation Computer Systems, 2018, 78964–975
DOI: 10.1016/j.future.2016.11.031

5 Stergiou C L, Psannis K E, Ishibashi Y. Green cloud communication system for big data management. In: 2020 3rd World Symposium on Communication Engineering (WSCE). Thessaloniki, Greece, IEEE, 2020, 69–73
DOI: 10.1109/wsce51339.2020.9275579

6 Xie F, Yan J, Shen J. A novel independent job rescheduling strategy for cloud resilience in the cloud environment. Applied Computing and Informatics, 2022
DOI: 10.1108/aci-06-2021-0172

7 Lin F, Dai W, Li W, Xu Z, Yuan L. A framework of priority-aware packet transmission scheduling in cluster-based industrial wireless sensor networks. IEEE Transactions on Industrial Informatics, 2020, 16(8): 5596–5606
DOI: 10.1109/tii.2019.2944980

8 Batalla J M, Mavromoustakis C X, Mastorakis G, Xiong N N, Wozniak J. Adaptive positioning systems based on multiple wireless interfaces for industrial IoT in harsh manufacturing environments. IEEE Journal on Selected Areas in Communications, 2020, 38(5): 899–914
DOI: 10.1109/jsac.2020.2980800

9 Batalla J M. On analyzing video transmission over wireless WiFi and 5G C-band in harsh IIoT environments. IEEE Access, 2020, 8118534–118541
DOI: 10.1109/access.2020.3005641

10 Granat J, Batalla J M, Mavromoustakis C X, Mastorakis G. Big data analytics for event detection in the IoT-multicriteria approach. IEEE Internet of Things Journal, 2020, 7(5): 4418–4430
DOI: 10.1109/jiot.2019.2957320

11 McMahan H B, Moore E, Ramage D, Hampson S, y Arcas B A. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017), MLR: W&CP, Fort Lauderdale, Florida, USA, 2017

12 Zhang W, Lu Q, Yu Q, Li Z, Liu Y, Lo S K, Chen S, Xu X, Zhu L. Blockchain-based federated learning for device failure detection in industrial IoT. IEEE Internet of Things Journal, 2021, 8(7): 5926–5937
DOI: 10.1109/jiot.2020.3032544

13 Nilsson A, Smith S, Ulm G, Gustavsson E, Jirstrand M. A performance evaluation of federated learning algorithms. In: Proceedings of the Second Workshop on Distributed Infrastructures for Deep Learning. Rennes France, New York, NY, USA/ACM, 2018, 1–8
DOI: 10.1145/3286490.3286559

14 Zhou J, Zhang S, Lu Q, Dai W, Chen M, Liu X, Pirttikangas S, Shi Y, Zhang W, Herrera-Viedma E. A survey on federated learning and its applications for accelerating industrial internet of things. 2021

15 Lv Z, Wu J, Li Y, Song H. Cross-layer optimization for industrial Internet of Things in real scene digital twins. IEEE Internet of Things Journal 2022, 1
DOI: 10.1109/jiot.2022.3152634

16 Aujla G S, Kumar N. MEnSuS: An efficient scheme for energy management with sustainability of cloud data centers in edge–cloud environment. Future Generation Computer Systems, 2018, 861279–1300
DOI: 10.1016/j.future.2017.09.066

17 Al-Dulaimy A, Itani W, Zantout R, Zekri A. Type-aware virtual machine management for energy efficient cloud data centers. Sustainable Computing: Informatics and Systems, 2018, 19185–203
DOI: 10.1016/j.suscom.2018.05.012

18 Khan M A, Paplinski A, Khan A M, Murshed M, Buyya R. Dynamic virtual machine consolidation algorithms for energy-efficient cloud resource management: a review. Sustainable Cloud and Energy Services, 2017, 6: 135–165
DOI: 10.1007/978-3-319-62238-5_6

19 Weber I, Wada H, Fekete A, Liu A, Bass L. Automatic undo for cloud management via AI planning. In: Proceedings of the Eighth USENIX

conference on Hot Topics in System Dependability (HotDep′12), USENIX, 2012
20 Brown L E, Kauchak D. Educational advances in artificial intelligence. AI Magazine, 2014, 34(4): 127
DOI: 10.1609/aimag.v34i4.2508
21 Rad P, Roopaei M, Beebe N, Shadaram M, Au Y. AI thinking for cloud education platform with personalized learning. In: Proceedings of the 51st Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences, 2018
DOI: 10.24251/hicss.2018.003
22 Wu C, Buyya R, Ramamohanarao K. Big Data Analytics=Machine Learning+Cloud Computing. 2016
23 Ahmed E, Yaqoob I, Hashem I A T, Khan I, Ahmed A I A, Imran M, Vasilakos A V. The role of big data analytics in Internet of Things. Computer Networks, 2017, 129459–471
DOI: 10.1016/j.comnet.2017.06.013
24 Lee J, Davari H, Singh J, Pandhare V. Industrial artificial intelligence for industry 4.0-based manufacturing systems. Manufacturing Letters, 2018, 1820–23
DOI: 10.1016/j.mfglet.2018.09.002
25 Wan J, Yang J, Wang Z, Hua Q. Artificial intelligence for cloud-assisted smart factory. IEEE Access, 2018, 655419–55430
DOI: 10.1109/access.2018.2871724
26 Khan S, Shakil K A, Alam M. Cloud-based big data analytics—A survey of current research and future directions. Springer, Big Data Analytics, 2017, 654: 595–604
DOI: 10.1007/978-981-10-6620-7_57
27 Lee J, Bagheri B, Kao H A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 2015, 318–23
DOI: 10.1016/j.mfglet.2014.12.001
28 Alhamad M, Dillon T, Chang E. Conceptual SLA framework for cloud computing. In: 4th IEEE International Conference on Digital Ecosystems and Technologies. Dubai, United Arab Emirates, IEEE, 2010, 606–610
DOI: 10.1109/dest.2010.5610586
29 Koutitas G, Demestichas P. Challenges for energy efficiency in local and regional data centers. Journal of Green Engineering, 2010, 1: 1–32
30 Manjunatha H., Esfahani E T. Application of reinforcement and deep learning techniques in brain–machine interfaces. Springer, Advances in Motor Neuroprostheses, 2010, 1–14
DOI: 10.1007/978-3-030-38740-2_1
31 Botvinick M, Ritter S, Wang J X, Kurth-Nelson Z, Blundell C, Hassabis D. Reinforcement learning, fast and slow. Trends in Cognitive Sciences, 2019, 23(5): 408–422
DOI: 10.1016/j.tics.2019.02.006