

Received April 13, 2020, accepted April 28, 2020, date of publication May 6, 2020, date of current version June 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992748

A Survey and Taxonomy on Energy-Aware Data Management Strategies in Cloud Environment

XINDONG YOU¹, XUEQIANG LV¹, ZHIKAI ZHAO², JUNMEI HAN³, AND XUEPING REN⁴

¹Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science & Technology University, Beijing 100101, China

²National Joint Engineering Laboratory of Internet Applied Technology of Mines, Internet of Things (IoT) Research Center, China University of Mining and Technology, Xuzhou 221008, China

³Laboratory of Complex Systems, Institute of Systems Engineering, AMS, PLA, Beijing 100102, China

⁴Key Laboratory of Complex Systems Modeling and Simulation, Ministry of Education, Hangzhou Dianzi University, Hangzhou 310037, China

Corresponding author: Xueqiang Lv (lxq@bistu.edu.cn)

This work is supported by National Natural Science Foundation of China under Grants No. 61671070, National Science Key Lab Fund project 6142006190301, National Language Committee of China under Grants ZD1135-53, and Project of and Project of Developing University Intension for Improving the Level of Scientific Research-No.No.2019KYNH226, Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University No. QXTCP B201908, Zhejiang Provincial Natural Science Foundation of China (NO. LQY18F020001).

ABSTRACT During the past ten years, the energy consumption problem in cloud-related environments has attracted substantial attention in research and industrial communities. Researchers have conducted many surveys on energy efficiency issues from different perspectives. All of the surveys can be classified into five categories: surveys on the energy efficiency of the whole cloud related system, surveys on the energy efficiency of a certain level or component of the cloud, surveys on all of the energy efficient strategies, surveys on a certain energy efficiency techniques, and other energy efficiency related surveys. However, to the best of our knowledge, surveys on energy-aware data management strategies in cloud-related environment are absent. In this paper, we conduct a comprehensive survey on energy saving-aware data management strategies in cloud-related environments, such as data classification, data placement and data replication strategies. Compared to current existing reviews on energy efficiency in cloud-related environments, we firstly conduct the survey on the energy consumption problem from the data management perspective. Furthermore, we classify the energy-aware data management strategies from different perspectives. This survey and the taxonomy of the energy-aware data management strategies demonstrate the potential for reducing the energy consumption at the data management level of a cloud storage system, which will compress more space for energy reduction and finally achieve energy proportionality. Moreover, this survey and taxonomy on the energy efficiency issue from the data management perspective is an important supplement to current existing surveys on energy efficiency in cloud-related environments.

INDEX TERMS Energy consumption, cloud storage system, data classification, data placement, data replication, energy proportionality.

I. INTRODUCTION

Energy consumption is one of the most serious problems in cloud-related environments, especially in datacenter part. Statistics indicate that the data volume is increasing at a rate of 50% every year, and the total of data volume is predicted to reach 40ZB in 2020 [1] (as shown in Figure 1). The exponential growth of the data volume leads to an increasingly more serious energy consumption problem. It has been reported by

The associate editor coordinating the review of this manuscript and approving it for publication was Honghao Gao¹.

literature [2] that the energy consumed by data centers will be more than 1000TWh in 2013-2025, which will surpass the total energy consumption of Japan and Germany. In addition, the energy consumption of the data centers and their cooling equipments will reach 5% of the total energy consumption of the world. Furthermore, the increasing energy consumption will produce high carbon and GHG (Greenhouse Gases) emissions [3], which will result in serious environmental pollution.

As previously mentioned, the serious energy consumption problem makes it important to improve energy efficiency

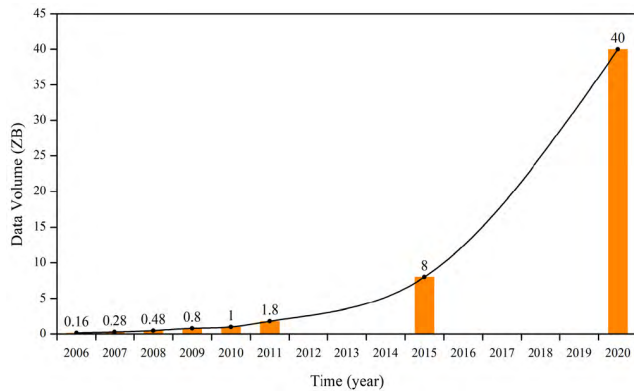


FIGURE 1. Data grow chart.

to reduce the operational costs of data centers and relieve environment pollution. During the past ten decades, many scholars have proposed or designed various strategies or algorithms from different perspectives.

From the macro perspectives, the energy efficiency techniques include energy-aware software techniques and energy efficient hardware techniques. Energy-aware software techniques include workload consolidation [4], [5], task or job scheduling [6]–[8], data concentration [9], data placement [10], data replication [11], [12], VM consolidation [13], [14], VM migration [15], [16] and VM scheduling [17], [18]. Energy efficient hardware techniques usually employ the energy efficient computing components [19], [20], such as, multispeed disks, SSD disks, DVFS enabled CPU and Flash memory etc. From the layer of cloud system perspective, energy efficient techniques are employed in different layers, such as, in application layer [21], data center layer [22], [23], cluster layer [24], [25], RAID layer [26], [27], node layer [28], virtualization layer [29], [30], OS layer [31], processor layer [32], [33], disk layer [34], [35], memory layer [36], [37] and network layer [38], [39]. The review map of the different energy-aware techniques is shown in Figure 2.

As shown in Figure 2, a large number of studies have focused on energy efficiency in cloud-related environments, which has made remarkable achievements with respect to energy consumption. However, the growing energy consumption is still an emergent issue in the cloud computing field. Many scholars have tried to search the new space to reduce the energy consumption by conducting surveys on energy consumption from different perspectives. In article [117], we classified the current existing surveys on energy consumption in cloud-related environment into five categories, which are, surveys on the energy efficiency of the whole cloud related system, surveys on energy efficiency of the certain level or component of the cloud, surveys on all of energy efficient strategies, surveys on a certain energy efficiency techniques, and other energy efficiency related survey. According to our observations, there is no article conducting the surveys and taxonomy of energy-aware data management strategies. Therefore, the survey and taxonomy of energy-aware data management strategies in cloud-related environments are

conducted in this paper, which is an important supplement for the current energy efficiency relevant surveys. Furthermore, the survey and taxonomy on improving energy efficiency utilizing the data management strategies, such as data classification, data placement and data replication, can offer more opportunities to reduce the energy consumption in cloud-related environments, especially for data-intensive related applications in the scale-growing data centers.

II. OUR PREVIOUS WORK AND FOCUS OF THIS SURVEY

As mentioned before, there are many related surveys on the energy efficiency in cloud-related environments over past the five years. We conducted our research from the survey perspective [117]: all of the surveys are classified into five categories: surveys on the energy efficiency of the whole cloud related system [40]–[50], surveys on the energy efficiency of the certain level or component of the cloud [39], [51]–[59], surveys on all energy efficient strategies [60]–[64], surveys on a certain energy efficiency technique [65]–[76], and other energy efficiency related surveys [77]–[80]. In addition, the different categories of surveys are summarized from five aspects, which include the title, survey focus, perspective, target system and publication years. From the summary and statistics on the existing surveys on the energy efficiency of cloud related environments, we have four observations: (1) energy efficiency surveys are related to many aspects of cloud related systems; (2) surveys on certain energy efficient strategies in cloud-related systems have aroused most concern; (3) the popularity of the research on energy efficiency issue on-going, and (4) surveys on energy-aware data management strategies in cloud related systems are absent.

Based on our previous observations, we focus our survey on the energy-aware data management strategies in this paper. The data management strategies in cloud related environments, usually include data classification strategies, data placement algorithms and data replication techniques. Therefore, we review the energy-aware data management strategies from the data classification, data placement and data replication perspectives, which aim to provide a new perspective and more space for improving the energy efficiency in cloud-related systems.

Our goal is to discover ways to optimize energy efficiency in cloud-related environments. To discover more opportunities to reduce the energy consumption in data centers, we focus on the lack of surveys on energy-aware data management strategies in cloud related environments. As a whole, the main contributions in this paper can be summarized as following items.

- 1) *The survey and taxonomy on energy saving strategies through employing data classification strategies, data layout policies, and data replication techniques from different perspectives are conducted in this paper.*
- 2) *The taxonomy and summary of the energy-aware data management strategies are done thereafter.*

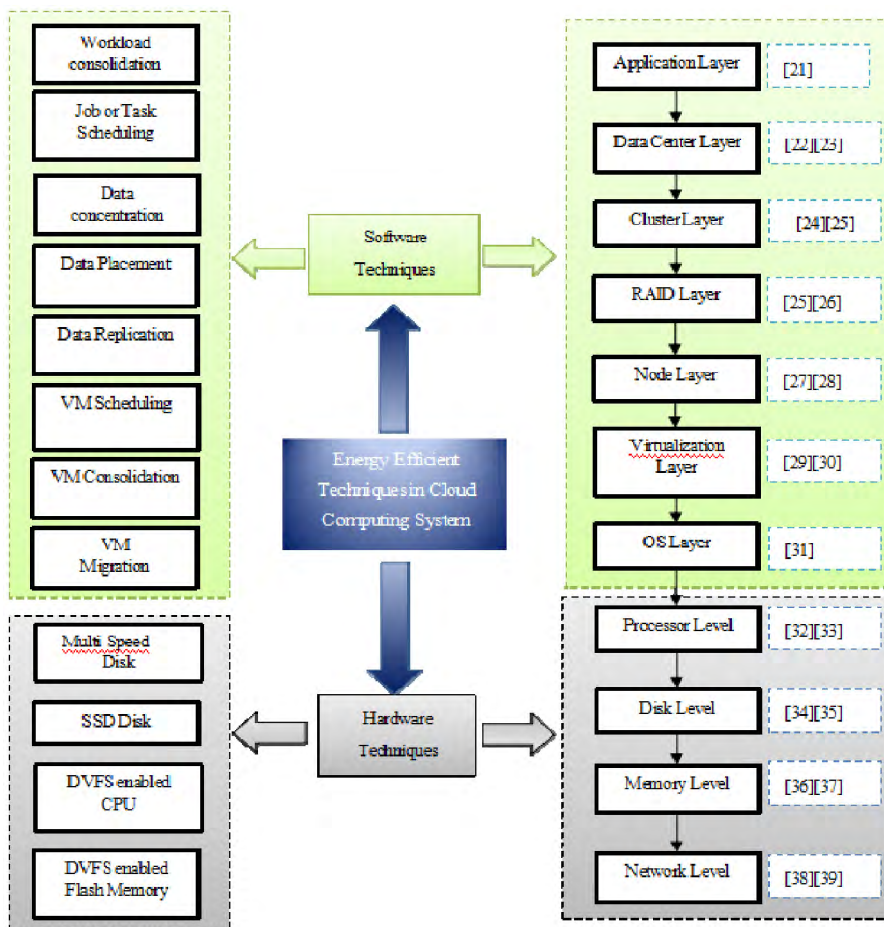


FIGURE 2. Review map of the different energy-aware technique.

- 3) Observations are deduced from the statistics on the taxonomy and the summary on energy-aware data strategies.
- 4) Finally, future directions for reducing energy consumption are comprehensively suggested.

The remaining parts of this paper are organized as follows. Energy saving strategies utilizing data classification, energy-aware data layout polices and energy savings strategies utilizing data replication techniques are described in section 3, section 4, and section 5 respectively. The Summary and observations of the energy-aware data management strategies are listed in section 6. We will conclude our paper and set forth our further work in the final section 7.

III. ENERGY SAVINGS STRATEGIES THROUGH DATA CLASSIFICATION

Energy-aware data classification strategies usually divide the storage system into different zones, and then the data are classified into different classes according to a certain rules. Therefore, different types of dataset are stored in their corresponding zones. Energy consumption savings are achieved by managing the power states of the different zones. Firstly,

T.Xie proposed a striping-based energy-aware strategy for data placement in RAID storage system [81], in which the storage system is divided into Hot Disk Zone and Cold Disk Zone. The Hot Disk Zone stores the popular data. The Cold Disk Zone stores the unpopular data. Disks in Hot Disk Zone run at high transfer rates and high power consume rates, while disks in Cold Disk Zone run at low transfer rates and low power rates. Analysis and simulation results show that the proposed SEA mechanism can noticeably reduce energy consumption with only a little performance degradation. Only mathematical analysis and simulation experiments have been conducted to evaluate the energy efficiency of the proposed SEA mechanism. According to the analysis of the Yahoo traces, the data access patterns in a Hadoop cluster are significantly heterogeneity. R.T.Kaushik et al design a GreenHDFS mechanism. GreenHDFS classified the data by their temperature, and the Hadoop cluster is divided into multi-zones (Hot Zone and Cold Zone). The data’s temperature is changed according to its availability or performance requirements. GreenHDFS mechanism utilizes the data’s heterogeneity and employs data classification techniques to place the data into the corresponding zone according to their temperature. The simulation results from three months of real trace from

Yahoo, show that only by managing the low power consumption rate of the Cold Zone, energy consumption can be reduced by 26% [82]. In the paper, GreenHDFS also has not been evaluated in real cloud environments, which means that its energy efficiency is not credible enough. Similar energy-aware data classification policies for a cloud storage system named Lighting were also designed by Kaushik team [83]. Inspired by the work of Kaushik's team, we have proposed a green data classification strategy based on anticipation named AGDC, in which a neural network is employed to predict the temperature of the data. Based on the predicted data's temperature, the data are classified into cold data, seasonal hot data and hot data. The cloud storage system is also divided into corresponding zones. Simulation experiments based on Gridsim show that the AGDC mechanism lowered energy consumption by 16% at the expense of increasing average response time by 0.005s. AGDC has its advantages while compared to the TDCS integrated general classification algorithm [84]. However, the temperatures predicted by the neural network will impact the energy efficiency of AGDC. In [88], the RACK is divided into an Active-Zone and a Sleep-Zone, and data are stored in the corresponding zone according to the data access regularity and frequency. Simulation results obtained from the MATLAB and Gridmix environment show that the proposed algorithm can save energy consumption by up to 39.01%. The performance degradation was not analyzed in the paper. In reference [85], an energy-efficient algorithm based on data classification for a cloud storage system is proposed by Z.Tao *et al.* They divided the cloud storage area into HotZone, ColdZone and Reduplication Zone. The data are stored in the corresponding zone based on the repetition and activity factor characteristics. The experimental results show that the proposed algorithm improves the energy utilization rate by nearly 25%. Furthermore, the algorithm performs well especially when the system load is light. However, three zones may induce frequent data migration, which will result in performance degradation. A dynamic data aggregation algorithm for green cloud computing is proposed in reference [86]. According to the data access pattern, the data and the nodes are aggregated and dynamically stored. By managing the power states of the storage nodes, the energy consumption can be reduced while considering QoS. There is the same problem as in reference [85]. Aiming to reduce the energy consumption in cloud storage systems, Dr. Long designed the static and dynamic file layout, replica and data layout policies [87]. The static file layout strategy (SFLS) first divided the data into hot files and big files according to their access frequency and service time, and the disks were correspondingly divided into different groups. The I/O requests were distributed to the different disk groups according to the access frequency and service time. The results obtained from the Cloudsim simulator demonstrate that the SFLS can save power consumption by over 35% while compared to the default HDFS. Evaluating Energy efficiency evaluated in real cloud environments is also absent in the paper. More recently, Yadav et al designed three adaptive energy-aware

algorithms to minimize the energy consumption and to reduce the SLA violations, in which the real workload traces are utilized to validate their feasibility [118]. However, the testing was limited to the simulation stage. Wang et al design a pipsCloud for the remote sensing of big data management and processing [119].

According to the above description, we summarize the taxonomy of the energy-aware data classification strategies from the aspects of the data classification criteria, the zones divisions, the experimental datasets, the experimental environment, the energy effectiveness and the publication year, which is shown in Table 1.

IV. ENERGY AWARE DATA LAYOUT POLICIES

In order to carry out a gear-shifting mechanism among the storage systems and to achieve power-proportionality, placing the data in a reasonable way is important. H.Amur designed a robust and flexible power-proportional storage named Rabbit [89]. Rabbit utilizes the equal-work data layout policy, which places the primary replica on the first ten nodes, the second replica is placed on the next ten nodes, and so on. The formulated policy and its implementation in the prototype Rabbit verified its power-proportionality. However, the data placement policy in Rabbit does not consider write requests when nodes are inactive. They evaluated the Rabbit and the PARAD's system performance costs for write access in low gear. The evaluated results showed that PARAD offers better performance when dealing with a frequently updated dataset [90]. Similar to Rabbit, Accordion, a data placement mechanism was proposed in literature [91], uses the elaborated data replication strategies to smooth gears shifting among the nodes. Substantial experiments conducted in the Hadoop DFS show that the Accordion mechanism can improve the power-proportional performance by 20% compared to the Rabbit Mechanism. N.Maheshwari proposed a dynamic energy efficient data placement and cluster reconfiguration algorithm for the MapReduce framework [92], in which nodes are turned on or off is according to the current workload and the extent to which the requirements are satisfied. Data are created or deleted to improve the performance or save power consumption while the nodes are turned on or off. Simulation experiments done on the Gridsim demonstrated the proposed algorithm can save 33% energy consumption under the average workload and up to 54% in the low workloads. And the experiments in real cloud environments are also absent. A semantic data placement algorithm designed for archival-by-accident workloads is described in reference [93]. They divided the data into access groups according to the semantic or incidental labels, including the file system placement, timestamps, the authors in a LaTeX document and file type etc. The grouped data ensure that the fast, consecutive accesses to the same group do not need an extra disk to spin-up, which can achieve power savings. Experiments from the California Department of Water Resources show that a 30% hit rate can result in at least 12% power savings. Similarly, R. Reddy et al designed a data layout for power

TABLE 1. Taxonomy of the energy-aware data classification strategies.

Literature	Data Classification Criterion	Zones Divided	Dataset in Experiment	Experimental Environment	Energy Effectiveness	Publication Year
SEA: A Striping-Based Energy-Aware Strategy for Data Placement in RAID-Structured Storage Systems [81]	Data's popularity	Hot Disk Zone and Cold Disk Zone	Synthetic workload	Disksim	Noticeably reduce energy consumption with only a little performance degradation	2008
GreenHDFS: towards an energy- conserving, storage-efficient, hybrid Hadoop compute cluster [82]	Data's spatial or temporal popularity	Cold Zone and Hot Zone	Three-month-long traces from Yahoo cluster	Hadoop Cluster at Yahoo	Average 26% Energy Reduction	2010
Dynamic data aggregation algorithm for data centers of green cloud computing [86]	Data's access pattern	Similar data access pattern aggregating stored	Synthesized workload	Simulated Data center	43.06% energy consumption conserved	2012
Energy-efficient algorithms for distributed storage system based on data storage structure reconfiguration [88]	Data access regularity and frequency.	Active-Zone and Sleep-Zone	Gridmix Benchmark in Hadoop	Matlab	Up to 39.01% energy consumption saved	2013
Research on Data Layout Strategies for Cloud Storage System [87]	Access frequency and service time	Hot file disks group and Big file disks group	Synthesized workload according to Zipf distribution	CloudSim Simulator	Over 35% energy consumption saved	2014
Energy-efficient algorithm based on data classification for cloud storage system [85]	Data's repetition and activity factor	HotZone, ColdZone and ReduplicationZone	Synthesized workload	CloudSim Simulator	Nearly 25% energy utilization improved	2014
Anticipation-based Green Data Classification Strategy in Cloud Storage System [84]	Predicted data's temperature	HotZone ColdZone Seasonal HotZone	Synthesized workload	CloudSim Simulator	Average 16% energy saved	2015
Adaptive Energy-Aware Algorithms for Minimizing Energy Consumption and SLA Violation in Cloud Computing [118]	Network traffic	VM Consolidation according to the status of the workload	Real workload traces	CloudSim Simulator	Average 12% energy saved and 22% SLA violation reduced	2018

efficient archival storage system, in which an access-aware intelligent data layout mechanism is provided [94]. A two-tier architecture that consist of online and offline disks is designed to store the archival data in the spin-down disks. The result obtained from experiments with real-world archival traces showed that the optimized data layout algorithm can achieve power savings up to 78% compared to the random data placement policy. Moreover, a Semi-RAID data layout policy based on a sequential data access pattern is proposed by X.Li et al [95], in which the grouping strategy is employed. The grouping strategy leaves only part of the whole array active and let the rest of the array in standby status. The analysis and experiments show that as to the typical video surveillance application, the proposed group strategy can achieve power consumption saving up to 28%. In recent years, energy-aware data placement algorithms have aroused the attention of Chinese scholars. In 2013, Y.W.Xiao et al integrated the data placement policy with the nodes scheduling strategy for energy savings [96]. A heuristic data placement policy and two node scheduling algorithms, which use the greedy algorithm to discover the plan to turn on minimum

nodes to cover the maximum data block were proposed. Simulation experiments conducted on Cloudsim showed that the proposed algorithm can save energy consumption under the constrained budget QoS requirement. Aiming at a heterogeneous Hadoop cluster, a snakelike data placement mechanism (SLDP) is proposed in literatures [97], [98]. SLDP first divide the storage nodes into virtual storage tiers (VST), and then circuitously place the data into the VST based on data's hotness, which can achieve the effective power control on the nodes according to their hotness to achieve energy savings. The experimental results from two real data-intensive applications demonstrate that the SLDP is energy efficient, saves spaces, and favors in heterogeneous Hadoop environment. An energy consumption optimization aware data placement layout is proposed by J.Song [99], in which the data are distributed to the nodes according to their processing abilities. The analysis and experiments performed on a modified Hadoop (LocalHadoop and Neo-Hadoop) environment demonstrated that the proposed data placement policy has a great advantage over the uniform Hash algorithm and has a narrow advantage over the uniform

Hash algorithm with stronger fairness. Dynamic adaptability to PowerCass is designed in literature [100], in which nodes are divided into three groups, they are, active, dormant and sleepy. Those three groups aim to respond to high, medium and low workloads respectively. In addition, the data are dynamically distributed among the groups according to the workload situations. Experiments conducted on the Apache Cassandra demonstrate that the energy savings can reach 66% when compared to the unmodified Cassandra. More recently, Song *et al.* designed a Modulo based on data placement algorithm to optimize the energy consumption in Mapreduce system [120]. In order to reduce the wasted energy, Modulo places data with the goals of “fairness of size”, “fairness of range”, and “best adaptability”, which are energy efficient without introducing additional costs and delaying data loading. However, the three algorithms are designed for MapReduce related applications, which mean that they can be adapted to the other kinds of applications. Due to the new resource management and allocation framework (YARN) in the HDFS system, the default data layout schemes are not energy efficient, literature [121] proposed a new data layout scheme, which exploits the heterogeneity of the computing resource characteristics. Servers are sorted by three sets (termed the high-performance set, the energy-efficient set and the inefficient set). Data blocks are placed in the high-performance set and energy-efficient set, and the replicas are placed in the energy-inefficient set. A comparison of experimental results shows that the new data layout scheme can significantly reduce the energy consumption at the slightly higher mean response time of the jobs. The new data layout scheme is also designed for the MapReduce framework related system, which also has no scalability.

As a whole, the taxonomy of the energy-aware data placement policies is summarized in Table 2.

V. ENERGY SAVINGS STRATEGIES UTILIZING DATA REPLICATION TECHNIQUE

Generally, a replication technique is utilized to assure the data availability, accelerate the data access speed, improve the workload balance and enhance the system performance. Recently, data replication techniques were also utilized to reduce the energy consumption in Cloud related environments. W.Lang et al first utilized the Chained Declustering (CD) replication strategy to conduct energy management in reference [101]. In this strategy, the replicas are placed in the CD ring, which enable the system to turn off some nodes and ensure data accessibility in light load conditions. Furthermore, they discuss and resolve the problem of the load balance among the remaining active nodes. Experiments conducted on a constructed system with 1000 nodes verify the energy efficiency of the proposed method. Jacob Leverich and Kozyrakis at Stanford University take advantage of the existing replicas on the Hadoop System, design a *covering subset*, which contains sufficient nodes, to ensure the data availability. Then the uncovered nodes can be set to inactive status to save energy consumption

during low server utilization. Experimental results that were obtained from the Hadoop cluster show that the fractional configuration among the nodes can save energy consumption by 9% to 50%, at the expense of slight system performance degradation. Replication as a tool to save energy consumption in RAID systems is investigated in literature [102]. They proposed a novel approach named iRGS, which employs the replication strategy to allow gradually gear shifting among the disks according to the varied workload. The applied replication strategy set super RAID Groups (RDGs) and ordinary RDGs to replicate data from each other at different rates to ensure that in every gear, the energy consumption and system performance can be traded off while assuring the data availability and users' requirements. A practical power-proportionality in the data center storage named Sierra is proposed by E. Thereska et al [103]. The goal of the layout of the replicas in Sierra is maintaining g available copies using only g/r servers, where r is the total number of replicas. The replicas are placed using power-aware grouping pattern, which relaxes the Naïve grouping constraint to some degree to achieve energy consumption savings, but a reasonable tradeoff between energy savings, rebuild parallelism and load balancing is introduced. A Distributed virtual log (DVL) is utilized to record the updates to the replicas that are powered down or failed, which ensures the write consistency. To evaluate the proposed Sierra, a full prototype is implemented using live traces from Hotmail servers. The experimental results demonstrated that the 23% of the energy consumption can be saved. An energy efficient replication mechanism based on node's addresses was proposed by Y.Y. Liu, in which the nodes are sequentially addressed by their racks and places. And the replicas in the nodes among the racks are not random but sequentially addressed in the same rack. The data replicas are placed in the nodes with the least number of replicas until it is full. Therefore, the access will be skewed in the subset of the nodes in the cloud storage system when the load is light. Powering down the nodes can achieve the energy consumption savings. In addition, the experiments that are conducted on the constructed Hadoop cluster verified its energy efficiency. By leveraging data access behavior, a power-aware data replication strategy is proposed in article [105]. According to the 80/20 rule (80% of the data access is often served by 20% of the data), they replicate small amounts of data that are frequently accessed in the hot nodes and set the nodes as always in the active state. The remaining 80% of data are placed in the code nodes that are in low power state. An access trace generated by the Zipf-distribution data access pattern, and the experiments performed on the simulator consisting of 16 data storage nodes and 1 metadata server demonstrate that the designed replication strategy is energy efficient. A replica management mechanism using the replication factor as one of the three phases for energy savings was designed by S.Q.Long et al [107]. Because the number of replicas influences the energy consumption, more replicas usually means more that storage space is occupied and more energy is consumed. The replica management

TABLE 2. Taxonomy of the energy-aware data placement policies.

Article	Main Idea of Data Layout	Test Data	Experimental Environment	Energy Improved	Publication Year
Robust and Flexible Power-Proportional Storage [89]	The replicas are placed in the continuous nodes	Micro benchmarks Terasort benchmark	Hadoop Infrastructure	Power-proportionality properties	2010
Efficient Gear-shifting for a Power-proportional Distributed Data-Placement Method [91]	Data layout with bisection method	Files distributed in the File System	Hadoop Distributed File System	power-proportionality with 20% performance improvement compared to Rabbit	2013
Dynamic Energy Efficient Data Placement and Cluster Reconfiguration Algorithm for MapReduce Framework [92]	Reconfigure or turning on or off the nodes according to the utilization	I/O request files varied with numbers and sizes	Gridsim Simulator	save 33% energy consumption under the average workload and up to 54% under the low workloads	2012
Semantic Data Placement for Power Management in Archival Storage [93]	Data are divided into access group according to semantic or incidental label	Two static file access traces	California Department of Water Resources	30% hit rate can result in a least 12% power savings	2015
Data layout for power efficient archival storage systems [94]	Store the archival data in the spin-down disks	Real-world archival traces	Prototype implemented by Python	78% power saving compared to the random data placement policy	2015
Semi-RAID: A Reliable Energy-aware RAID Data Layout for Sequential Data Access [95]	Data grouping makes only the part of Array active	Files distributed in the RAID	Disksim simulator	28% energy saving under the typical video surveillance application	2011
Energy Efficient Data Placement Algorithm and Node-Scheduling Policies in Cloud Computing Systems [96]	Use the greedy algorithm to discover the plan to turn on minimum nodes to cover maximum data block	Cloudlet synthetic data and TPC-H, Spawner generated data	CloudSim simulator	Energy Saved with the constraint budget QoS requirement	2013
SLDP: A Novel Data Placement Strategy for Large-Scale Heterogeneous Hadoop Cluster [97]	Circuitously place the data into the Virtual Storage Tiers based on data's hotness	Files distributed in the File System	Hadoop Cluster	Energy efficient with space saving under heterogeneous environment	2014
Energy Consumption Optimization Data Placement Algorithm for MapReduce System. [99]	Distribute Data to the node according to the processing ability.	WordCount, Terasort, MRBench	Modified Hadoop (LocalHadoop and NeoHadoop) environment	Great advantage over the uniform Hash algorithm and fine advantage over the uniform Hash algorithm with stronger fairness	2015
Dyn-PowerCass: Energy Efficient Distributed Store Based on Dynamic Data Placement Strategy [100]	Relocate data in the active, dormant and sleepy node groups	Generated workload according to the access pattern	Apache Cassandra	Up to 66% energy saved compared to unmodified Cassandra	2015
Modulo Based Data Placement Algorithm for Energy Consumption Optimization of MapReduce System [120]	Try to achieve the goals of "fairness of size", "fairness of range", and "best adaptability"	Benchmark	MapReduce framework related systems	Energy consumption of WordCount, Sort, MRBench can be reduced by 10.9%, 8.3% and 17% respectively, and time consumption is reduced by 7%, 6.3% and 7% respectively	2018
A New Data Layout Scheme for Energy-Efficient MapReduce Processing Tasks[121]	Place data blocks in the high-performance server set and energy-efficient server set, and place replicas in the energy-inefficient set.	Simulator in C language with statistical module	MapReduce framework related systems	Significantly reduce the energy consumption with slight increase in the mean response time of jobs.	2018

strategy aimed to minimize number of replicas. When the number of replicas in the cloud storage system is more than the minimum number, the replica deletion policy is carried out according to the throughput of the data nodes of the

replicas. Experimental results from conducted on Cloudsim simulator show that the proposed algorithm has advantage over the existing schemes in energy consumption. They also proposed a replication management strategy to optimize

the multi-objectives (containing of the energy consumption index) of cloud storage cluster named MORM. The replication factor and layout are improved by the artificial immune algorithms. The proposed MORM mechanism try to find out the near optimal solutions to balance the trade-offs among the file availability, load variance, mean service time, access latency and energy consumption. Substantive experiments done on the constructed Hadoop cluster showed that the proposed MORM mechanism outperforms the default replication management mechanism in Hadoop in terms of the load balance and response time. X.L.Cui et al dealt with the energy and fault-tolerance problems using shadow replication [109], in which the main process is associated with a suite of shadow processes. In addition, a profit-based optimization model is utilized to determine the optimal speed of the task in order to reduce energy consumption while maximizing profits. Experiments on the self-developed evaluation framework with three different benchmarks verified the energy efficiency of the shadow replication. According to the users' visiting characteristic, a dynamic energy-aware replication management strategy was investigated by Z.Y. Wang et al [110]. The main idea of the proposed strategy is transforming the users' access characteristics so that they can be used to compute the access hotness in a Block. According to the integrated hotness, when the hotness of DataNode n in a Hadoop cluster is under a certain threshold, it is put to sleep. Then, the Data in the sleepy DataNode is temporarily replicated on the emergent DataNode. Benefitting from the considerable number of sleepy nodes while the system workload is light, energy consumption savings can be achieved. Energy efficiency and network consumption are combined in the replication mechanism in [111]. In this research, the data are replicated closer to the data consumers, which may be a promising solution to minimize the bandwidth usage and network delays and save the energy consumption. The RM (Replication Management) model that is located in Datacenter DB computes the update and access rates in previous intervals and predicts the future values. Then the replication decision is done based on the predicted values, which aims to save energy and network consumption. Experiments that were conducted on the self-developed GreenCloud Simulator demonstrate its effectiveness with respect to energy consumption and network usage. Based on the consistent hashing distribution, the power-proportional replication mechanism named GreenCHT is proposed in [12]. In this approach, the replicas are organized in virtual tiers. The first replica of the object is placed in tier 0, and the second replica is placed in tier 1, the third replica is placed in tier 2, and so on. A power-mode predictive model is employed to predict the load of the next period, which can determine the power mode state and determine which tier should be active to handle the load and which tier should be powered down to achieve the energy consumption savings. In Addition, the log-replicas are designed to address writing consistency problem. A trace driven by twelve real enterprise data center workloads were collected from Microsoft Cambridge servers. The prototype that was

implemented in the Sleepdog demonstrates that the power-proportional replication mechanism can reduce the energy consumption up to 31%-60% under the different workload at the expense of a 4-5 ms higher response time. However, the replication management mechanism in GreenCHT does not consider the heterogeneity of the object, and all of the objects have the same number of replicas. In actual applications, the files in the cloud storage system are heterogeneous with respect to certain properties, especially with respect to the files' popularity. Hot files need more replicas to serve the requests, and cold files need fewer replicas to save energy and storage space. According to the heterogeneity of the files, an energy-aware adaptive file replication mechanism for data intensive systems named EAFR is designed in [113], in which the number of the replicas of the data is decided by its hotness. The replica selection is also according to the heterogeneity of the server. The server with more capacity is selected first. Moreover, the hot files stored on the hot servers run at high power rates to achieve quick response, and the cold data stored on the cold servers run at low power rates to save energy consumption. Experiments using the trace that were conducted on the Palmetto Cluster of Clemson University's demonstrate that more than 150kWh per day energy consumption can be saved in a cluster consisting of 300 servers. The approach based and inspired by the energy-efficient replicas placement strategy named Superset was proposed by X.Y.Luo in her PhD thesis [106]. Z.L. Shi designed an energy-aware replica management strategy, which includes the replica factor decision, the replica selection strategy and the replica placement algorithm [113]. A file's hotness is computed according to the life cycle characteristics and the access rate. In addition, the number of replicas is determined by the hotness of the file. Different degrees of file hotness cause the different numbers of replicas. Then, the file sets and node sets are divided according to the files' hotness. Files are organized using super sets, which can assure that different numbers of the replicas are stored in the different node sets. The replica placement strategy provides support to the power-proportional gear-shifting mechanism while guaranteeing the file's availability. Experiments conducted on the Cloudsim simulator showed that the proposed management strategy can lower energy consumption by up to 16% over the non-power-aware replica strategy. In 2015, D. Kliazovich et al briefly analyze an energy-aware replication management strategy [114]. In 2016, a comprehensive survey on the data replication techniques in cloud storage systems is discussed in [115], and it points out that energy-aware replication is one of the future directions of replication strategies. Recently, article [122] formulated the replication problem as an optimization problem, and it used a hybrid metaheuristic algorithm that combined the global search capability of the Particle Swarm Optimization (PSO) algorithm and the local search capability of the Tabu Search (TS) to achieve high-quality solutions. Simulation experiments conducted in MATLAB indicated that the proposed method outperforms other optimization algorithms in terms of energy

consumption and costs. However, experiments on the real cloud environment are lacking, which lowers the confidence in the proposed. An energy-aware and adaptive fog storage mechanism that uses spatio-temporal content popularity was designed in literature [125]. In the mechanism, the factors of user data demand, energy consumption and node distance are considered to determine to replicate data to the node. Testbed results verified the energy efficiency and adaptability of this approach. However, the storage, processing and network features have not been considered in the current work. There are some related work can be used as references [124]–[134], which are not described in our body text.

Taxonomy of the energy savings strategies utilizing data replication technique is summarized in Table 3.

VI. SUMMARY AND OBSERVATIONS OF THE ENERGY-AWARE DATA MANAGEMENT STRATEGIES

The above surveys on energy-aware data management strategies demonstrate that it is possible to reduce energy consumption through elaborately designed data management strategies. Furthermore, the energy-aware data management strategies provide more space and opportunities to reduce energy consumption in the cloud related environment, which is important for supplementing other energy saving techniques, such as resource allocation, workload consolidation, VM scheduling, VM migration, VM consolidation and workload characterization. According to the above investigation, the energy-aware data management strategies usually include three categories, which are data classification, data layout and data replication techniques. In addition, we have the following observations.

Observation 1: Data replication is the main technique of the energy-aware data management strategies

To our best knowledge, all of the published articles about energy-aware data management strategies are usually fall into three categories, data classification, data layout and data replication techniques. The numbers of the three categories are shown in the Figure 3.

As shown in Figure 3, among the energy-aware data management strategies, the proportion of data classification mechanisms 27%, the data layout polices is 29%, and the data replication techniques is 44%. Obviously, most of the studies are conducted on the energy-aware data replication techniques, which imply that data replication is the most important technique in energy-aware data management field, especially for solving power-proportionality problems.

Observation 2: Energy-aware data management strategies have been comprehensively and thoroughly investigated. Since the concept of the cloud originated in 2008, energy-aware data management strategies have also generated aroused concern. Scholars have continued investigating this field until it reach the peak in 2015. We searched every possible keyword about energy and data management to find the related published papers. Unfortunately, there are no published papers on the topic in two nearest years 2016 and 2017. Recently, there are 5 papers published in 2018, and

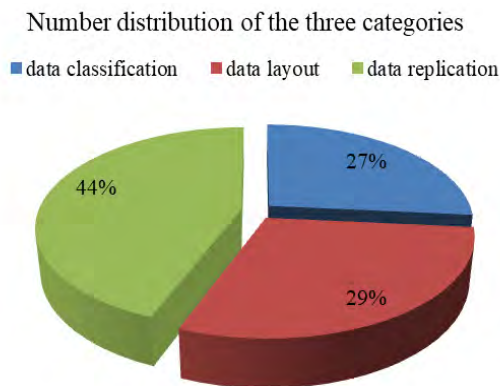


FIGURE 3. Number distribution of the three categories of the energy-aware data management strategies.

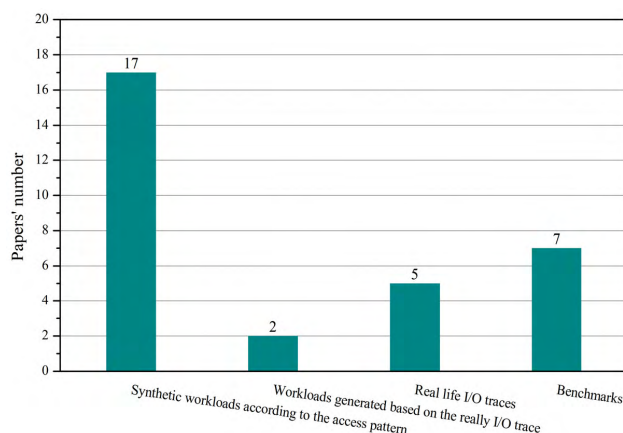


FIGURE 4. Numbers of papers published in past ten years.

1 paper published in 2019. Accordingly, we speculate that the research on energy consumption utilizing data management strategies has matured, and the related methods have been comprehensively and thoroughly investigated. Furthermore, the detailed numbers of papers that have been published during the past ten years in Figure 4 verified this speculation. Innovative research methods should to be excavated in the future.

Observation 3: Workload that are adopted while evaluating the methods usually fall into four categories. With respect to evaluating the energy efficiency of data management strategies, the workloads that are adopted usually fall into four categories, which are synthetic workloads according to the access pattern, and workloads that are generated based on the real I/O traces, real life I/O traces, and benchmarks. The numbers in the four categories are exhibited in Figure 5. Among these categories, synthetic workloads and workloads based on benchmarks are utilized more often than the other two categories.

Observation 4: Experimental environments usually fall into four categories. In order to evaluate the energy efficiency of the proposed data management strategies, the experimental environments that are utilized usually fall into four categories, they are prototype system, simulator or simulator extension,

TABLE 3. Taxonomy of the energy savings strategies utilizing data replication technique.

Article	Main Idea of Data Layout	Test Data	Experimental Environment	Energy Improved	Publication Year
On Energy Management, Load Balancing and Replication [101]	Replicas placed in the Chained Declustering ring	Wisconsin Benchmark Query 3 and file scan on a WB table	System constructed with 1000 nodes	Trade off load balancing for the energy efficiency	2009
On the energy (in)efficiency of Hadoop clusters [24]	Covering subset containing nodes to ensure data availability	Webdata_sort and webdata_scan	Hadoop environment	18 disabled nodes achieved 44% energy saved	2010
Using Replication for Energy Conservation in RAID Systems [102]	Different RAID groups replicate different shares of the others data	Real life trace data (Cello 99)	DiskSim Simulator	Energy saved up to 60, and out performs existing energy saving algorithm	2010
Sierra: practical power-proportionality for data center storage [103]	Power-aware grouping replica pattern that relax the Naïve random grouping is designed	I/O traces from Hotmail	Full prototype of the proposed mechanism	Significant energy consumption saved with little performance loss	2011
Research on Energy Efficient Replication Management and Task Scheduling under Cloud Environment [104]	Replicate the data in the incremental addressing nodes	WordCount Sort TeraSort Benchmark	Hadoop Clusters self-constructed	In the WordCount and Sort workload the nodes' replicating energy model has advantage the kernel zone model	2011
Designing a Power-aware Replication Strategy for Storage Clusters [105]	Replicating the frequently accessed data in the active nodes, and place the infrequently accessed data in low power state nodes	Zipf-like distribution File Access pattern Generated	Simulator constructed that consists of 16 storage nodes and 1 metadata server	Significant power reduction with performance guaranteed	2013
Research on Replica Placement Algorithms towards High-Efficient Clustered Storage Systems [106]	Files super set in corresponding with the nodes set to support the energy gear-shifting mechanism	Zipf-like file access collected from Youku and ACM	Self-constructed large-scale cluster storage system	Reduce the energy cost over by 50% at the same of I/O throughput	2013
A three-phase energy-saving strategy for cloud storage systems [107]	Keep minimum number of the replicas in the cloud storage system	Synthetic workload generator	Cloudsim simulator	Energy consumption can be reduced to approximately 40%	2014
MORM: A Multi-objective Optimized Replication Management Strategy for Cloud Storage Cluster [108]	Replication factor and layout are improved by the artificial immune algorithms	Synthetic workload generator	Cloudsim simulator	Outperforms the default replication management mechanism in Hadoop in terms of load the balance and performance.	2014
Shadow Replication: An Energy-Aware, Fault-Tolerant Computational Model for Green Cloud Computing [109]	Utilize the profit-based optimization model to decide the optimal speed of the task as to reduce energy consumption while maximizing profits	Three benchmark applications: Business Intelligence, Bioinformatics and Recommendation System	Self-developed evaluation framework	Up to 30% higher profit due to the energy consumption savings	2014
Energy-efficient strategy for dynamic management of cloud storage replica based on user visiting characteristic [110]	Temporarily replicate the data of the sleepy nodes on the emergent nodes	Synthetic workload generator	Self-constructed Hadoop cluster	Powering down 29-42% DataNode with 31% energy consumption saved	2014
Energy-efficient data replication in cloud computing datacenters [111]	Replicating the data closer to the data consumer using the RM model	Synthetic workload generated according to exponential distributed access pattern	GreenCloud Simulator	Energy efficiency and performance can be traded off	2015
GreenCHT: A Power-Proportional Replication Scheme for Consistent Hashing based Key Value Storage System [12]	Replicate the data in the different virtual tiers, and tiers are powered up or down according to the status of workload	Twelve real enterprise data center workload collected from Microsoft Cambridge servers	Sleepdog	Power consumption reduced by up to 35%-61%	2015

TABLE 3. (Continued.) Taxonomy of the energy savings strategies utilizing data replication technique.

EAFR: An Energy-Efficient Adaptive File Replication System In Data-Intensive Clusters [112]	The number of the replicas decided by the hotness considering the performance and energy consumption	Trace from HPC Palmetto Cluster of Clemson University's	Scattered cluster with 300 servers.	More than energy consumption saved in a cluster consisting of 300 servers 150kWh per day	2015
Research on Efficient Replica Management Strategy in Cloud Environment [113]	Files super set in corresponding with the nodes set to support the energy gear-shifting mechanism	Synthetic workload generated according to the life cycle characteristic	Cloudsim Simulator	Saving the energy consumption up to 16%	2015
An energy-aware method for data replication in the cloud environments using a Tabu search and particle swarm optimization algorithm [122]	Transform the replication problem into an optimization problem, and then apply the related algorithms to solve it.	Dataset generated according to the cloud system architecture	MATLAB	Achieve better performance in terms of energy and costs than the other optimization algorithms	2018
Energy-aware and adaptive fog storage mechanism with data replication ruled by spatio-temporal content popularity [128]	Replicating data according to spatial and temporal popularity	Synthetic workload generated according to the files' popularity	Testbed	Energy efficiency and framework with adaptability	2019

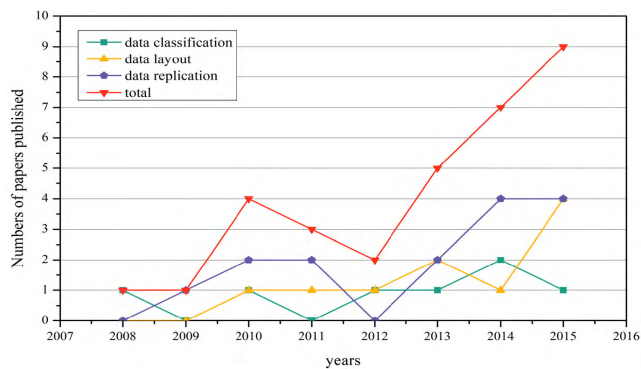


FIGURE 5. Numbers distribution on the four common workloads during evaluation.

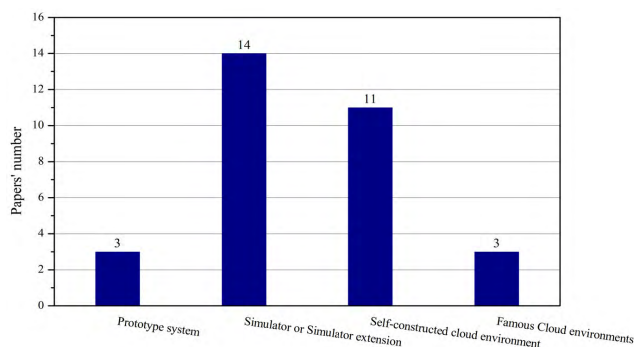


FIGURE 6. Different category of evaluation environment employed in the papers.

self-constructed cloud environment and actual business cloud environment. The numbers of papers utilized in the three categories of involving experimental environments are provided in Figure 6.

As shown in Figure 6, the most common environments that are employed in evaluation experiments are simulators and self-constructed cloud environments. The use of

self-designed prototype systems is limited due to the mass of developing work that are required for these systems. On the other hand, utilizing famous actual business cloud environments is a very small part due to the mass of debugging work.

VII. CONCLUSION AND FUTURE DIRECTIONS

Based on our previous work and observations on the survey and taxonomy of the current existing surveys on energy efficiency strategies in cloud related environments, we focus this survey on energy-aware data management strategies. Energy-aware data management strategies are classified into three categories, which are *data classification strategies*, *data layout policies* and *data replication techniques*. For every data management category, we investigate its main ideas, test data, experimental environments and published years respectively. Finally, the observations during the reviewing of energy-aware data management strategies are presented.

There are three research directions for energy saving techniques in our future work.

Future direction 1: Reducing energy consumption through multi-level combination

As energy consumption savings strategies in cloud related environments has been attracted comprehensive investigations over the past ten years, there is little space to reduce energy consumption at single level of cloud related systems. Furthermore, the respective energy efficient strategies may conflict with each other when integrating them into real cloud systems, which make the designed energy consumption strategies inefficient. Therefore, how to combine the energy efficiency strategies in every level and to form a holistic energy efficient framework is one of the future directions.

Future direction 2: Combine data classification, data layout and data replication techniques to pursue further energy saving e

Based on our investigation on the energy-aware data management strategies in cloud related systems, the one-fold data management strategy, such as only utilizing data classification, data layout or data replication technique, has been thoroughly studied, and effective energy savings have been achieved. Combining different data management strategies into an integrated energy efficient framework may achieve more energy saving space, which is one of the future research directions in the energy efficiency field.

Future direction 3: Combine energy-aware data management strategies with other traditional energy efficient techniques.

Data management is an important supplementary technique for energy efficiency in cloud related environments, since energy-aware data management and traditional energy efficient strategies have been thoroughly investigated. Combining data layout policies and data replication with energy-aware scheduling algorithms, DVFS techniques and energy gear-shifting mechanism, may be one of the future directions for further improving energy efficiency in cloud related systems.

REFERENCES

- [1] *The Developing Trends and Market Scale of BigData Industry in China in 2016 [EB/OL]*. Accessed: Apr. 29, 2016. [Online]. Available: http://www.cnii.com.cn/Bigdata/2016-04/29/content_1723221.htm
- [2] M. P. Mills. *The Cloud Begins With Coal: Big Data, Big Networks, Big Infrastructure, and Big Power [EB/OL]*. Accessed: Apr. 25, 2013. [Online]. Available: http://www.tech-pundit.com/wpcontent/uploads/2013/07/Cloud_Begins_With_Coal.pdf
- [3] *New Report Reveals Warehouses Overspend on Energy by \$190 m*, Data World Resour. Inst., Envanantage, Cleveland, OH, USA, Aug. 2013.
- [4] S. Srikantaiah, A. Kansal, and F. Zhao, "Energy aware consolidation for cloud computing," in *Proc. Workshop Power Aware Comput. Syst. (HotPower)*, San Diego, CA, USA, 2008, pp. 1–5.
- [5] G. Prekas, M. Primorac, A. Belay, C. Kozyrakis, and E. Bugnion, "Energy proportionality and workload consolidation for latency-critical applications," in *Proc. 6th ACM Symp. Cloud Comput. (SoCC)*, Kohala Coast, HI, USA, Aug. 2015.
- [6] Q. Tang, S. K. S. Gupta, and G. Varsamopoulos, "Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 11, pp. 1458–1472, Nov. 2008.
- [7] T. V. T. Duy, Y. Sato, and Y. Inoguchi, "Performance evaluation of a green scheduling algorithm for energy savings in cloud computing," in *Proc. IEEE Int. Symp. Parallel Distrib. Process., Workshops Phd Forum (IPDPSW)*, Apr. 2010, pp. 1–8.
- [8] H. Chen, X. Zhu, H. Guo, J. Zhu, X. Qin, and J. Wu, "Towards energy-efficient scheduling for real-time tasks under uncertain cloud computing environment," *J. Syst. Softw.*, vol. 99, pp. 20–35, Jan. 2015.
- [9] D. K. Lee and K. Koh, "PDC-NH: Popular data concentration on NAND flash and hard disk drive," in *Proc. 10th IEEE/ACM Int. Conf. Grid Comput.*, Oct. 2009, pp. 196–200.
- [10] R. Reddy, A. Kthpal, and J. Basak. (2015). Data Layout for Power Efficient Archival Storage Systems. HotPower [Online]. Available: https://engineering.purdue.edu/~hotpower/HotPower_2015_paper_2.pdf
- [11] S. Iwata and K. Shiozawa, "A simulation result of replicating data with another layout for reducing media exchange of cold storage," in *Proc. 8th USENIX Workshop Hot Topics Storage File Syst. (HotStorage)*, Denver, CO, USA, Jun. 2016, pp. 1–5.
- [12] N. Zhao, J. Wan, J. Wang, and C. Xie, "GreenCHT: A power-proportional replication scheme for consistent hashing based key value storage systems," in *Proc. 31st Symp. Mass Storage Syst. Technol. (MSSST)*, Santa Clara, CA, USA, May/June. 2015, pp. 1–6. [Online]. Available: <http://storageconference.us/2015/Papers/16.Zhao.pdf>
- [13] F. Farahnakian, A. Ashraf, P. Liljeberg, T. Pahikkala, J. Plosila, I. Porres, and H. Tenhunen, "Energy-aware dynamic VM consolidation in cloud data centers using ant colony system," in *Proc. IEEE 7th Int. Conf. Cloud Comput.*, Jun./Jul. 2014, pp. 104–111.
- [14] D. K. Kang, F. Alhazemi, S. H. Kim, and C. H. Youn, "Dynamic virtual machine consolidation for energy efficient cloud data centers," in *Proc. Int. Conf. Cloud Comput. (CloudComp)*, 2015, pp. 70–80.
- [15] M. H. A. Shayeji and M. D. Samrajesh, "An energy-aware virtual machine migration algorithm," in *Proc. Int. Conf. Adv. Comput. Commun.*, Aug. 2012, pp. 242–246.
- [16] V. R. Reguri, S. Kogatam, and M. Moh, "Energy efficient traffic-aware virtual machine migration in green cloud data centers," in *Proc. IEEE IEEE 2nd Int. Conf. Big Data Secur. Cloud (BigDataSecurity) Int. Conf. High Perform. Smart Comput. (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS)*, Apr. 2016, pp. 268–273.
- [17] C. Ghribi, M. Hadji, and D. Zeghlache, "Energy efficient VM scheduling for cloud data centers: Exact allocation and migration algorithms," in *Proc. 13th IEEE/ACM Int. Symp. Cluster, Cloud, Grid Comput.*, May 2013, pp. 671–678.
- [18] Y. Wang and Y. Xia, "Energy optimal VM placement in the cloud," in *Proc. IEEE 9th Int. Conf. Cloud Comput. (CLOUD)*, Jun./Jul. 2016, pp. 84–91.
- [19] A. Shafiee, A. Naq, and N. Muralimanohar, "ISAAC: A convolutional neural network accelerator with *in-situ* analog arithmetic in crossbars," in *Proc. 43rd Int. Symp. Comput. Archit. (ISCA)*, 2016, pp. 14–26.
- [20] M. M. Ozdal, S. Yesil, T. Kim, A. Ayupov, J. Greth, S. Burns, and O. Ozturk, "Energy efficient architecture for graph analytics accelerators," in *Proc. ACM/IEEE 43rd Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 2016, pp. 166–177. [Online]. Available: <https://www.cs.utah.edu/~rajeev/pubs/isca16.pdf>
- [21] W. Wang and K. Sohraby, "A study of application layer paradigm for lower layer energy saving potentials in cloud-edge social user wireless image sharing," in *Proc. 8th Int. Conf. Mobile Multimedia Commun.*, May 2015, pp. 54–59.
- [22] L. Zhou, C. Dong, X. You, J. Huang, and C. Jiang, "High availability green gear-shifting mechanism in cloud storage system," *Int. J. Grid Distrib. Comput.*, vol. 8, no. 5, pp. 303–314, Oct. 2015, doi: 10.14257/Ijgcd.2015.8.5.30.
- [23] Q. Wu, Q. Deng, L. Ganesh, C.-H. Hsu, Y. Jin, S. Kumar, B. Li, J. Meza, and Y. J. Song, "Dynamo: Facebook's data center-wide power management system," in *Proc. ACM/IEEE 43rd Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 2016, pp. 469–480.
- [24] J. Leverich and C. Kozyrakis, "On the energy (in)efficiency of Hadoop clusters," *ACM SIGOPS Operating Syst. Rev.*, vol. 44, no. 1, pp. 61–65, Mar. 2010.
- [25] V. Vasudevan, D. Andersen, M. Kaminsky, L. Tan, J. Franklin, and I. Moraru, "Energy-efficient cluster computing with FAWN: Workloads and implications," in *Proc. 1st Int. Conf. Energy-Efficient Comput. Netw. (e-Energy)*, Apr. 2010, pp. 195–204.
- [26] C. Weddle, M. Oldham, J. Qian, and A.-I. A. Wang, "PARAID: A gear-shifting power-aware RAID," in *Proc. 5th USENIX Conf. File Storage Technol. (FAST)*, Berkeley, CA, USA: USENIX, 2007, pp. 245–260.
- [27] E. Li and J. Wang, "EERAID: Energy efficient redundant and inexpensive disk array," in *Proc. 11th ACM SIGOPS Eur. Workshop*. New York, NY, USA: ACM, 2004, p. 29-es.
- [28] J. Kim and D. Rotem, "Energy proportionality for disk storage using replication," in *Proc. 14th Int. Conf. Extending Database Technol. (EDBT/ICDT)*. New York, NY, USA: ACM, 2011, pp. 81–92.
- [29] J. Stoess, C. Lang, and F. Bellosa, "Energy management for hypervisor-based virtual machines," in *Proc. USENIX Annu. Tech. Conf.* Santa Clara, CA, USA: USENIX Association, 2007, pp. 1–14.
- [30] B. Riskhan, K. Zhou, and R. Muhammad, "Energy management of the system: An empirical investigation of virtualization approaches in static and dynamic modes," *Inf. Technol. J.*, vol. 16, no. 1, pp. 1–10, Dec. 2017.
- [31] G. Y. Jia, J. J. Wan, X. Li, C. F. Jiang, and D. Dai, "Memory power optimization policy of coordinating page allocation and group scheduling," (in Chinese), *Ruan Jian Xue Bao/J. Softw.*, vol. 25, no. 7, pp. 1403–1415, 2014. [Online]. Available: <http://www.jos.org.cn/1000-9825/4600.htm>
- [32] C.-H. Hsu and U. Kremer, "The design, implementation, and evaluation of a compiler algorithm for CPU energy reduction," in *Proc. ACM SIGPLAN Conf. Program. Lang. Design Implement. (PLDI)* May 2003, pp. 38–48.

- [33] K. Ma, X. Li, W. Chen, C. Zhang, and X. Wang, "GreenGPU: A holistic approach to energy efficiency in GPU-CPU heterogeneous architectures," in *Proc. 41st Int. Conf. Parallel Process.*, Sep. 2012, pp. 48–57.
- [34] S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke, "DRPM: Dynamic speed control for power management in server class disks," in *Proc. 30th Annu. Int. Symp. Comput. Archit.* New York, NY, USA: ACM, 2003, pp. 169–181.
- [35] E. A. Papathanasiou and L. M. Scott, "Energy efficient prefetching and caching," in *Proc. USENIX Annu. Tech. Conf. (USENIX)*. Berkeley, CA, USA: USENIX, 2004, pp. 255–268.
- [36] M. E. Tolentino, J. Turner, and K. W. Cameron, "Memory MISER: Improving main memory energy efficiency in servers," *IEEE Trans. Comput.*, vol. 58, no. 3, pp. 336–350, Mar. 2009.
- [37] J. S. Vetter and S. Mittal, "Opportunities for nonvolatile memory systems in extreme-scale high-performance computing," *Comput. Sci. Eng.*, vol. 17, no. 2, pp. 73–82, Mar. 2015.
- [38] K. Cengiz and T. Dag, "A review on the recent energy-efficient approaches for the Internet protocol stack," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 1, pp. 1–22, Dec. 2015.
- [39] C. Eyupoglu and M. A. Aydin, "Energy efficiency in backbone networks," *Procedia-Social Behav. Sci.*, vol. 195, pp. 1966–1970, Jul. 2015.
- [40] A. Beloglazov, R. Buyya, Y. C. Lee, and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems," *Adv. Comput.*, vol. 82, pp. 47–111, Jan. 2011.
- [41] G. L. Valentini, W. Lassonde, S. U. Khan, N. Min-Allah, S. A. Madani, J. Li, L. Zhang, L. Wang, N. Ghani, J. Kolodziej, H. Li, A. Y. Zomaya, C.-Z. Xu, P. Balaji, A. Vishnu, F. Pinel, J. E. Pecero, D. Kliazovich, and P. Bouvry, "An overview of energy efficiency techniques in cluster computing systems," *Cluster Comput.*, vol. 16, no. 1, pp. 3–15, Mar. 2013.
- [42] A.-C. Orgerie, M. D. D. Assuncao, and L. Lefevre, "A survey on techniques for improving the energy efficiency of large-scale distributed systems," *ACM Comput. Surv.*, vol. 46, no. 4, pp. 1–31, Apr. 2014.
- [43] T. Kaur and I. Chana, "Energy efficiency techniques in cloud computing: A survey and taxonomy," *ACM Comput. Surveys*, vol. 48, no. 2, pp. 1–46, Nov. 2015.
- [44] S. Sohrabi and I. Moser, "A survey on energy-aware cloud," *Eur. J. Adv. Eng. Technol.*, vol. 2, no. 2, pp. 80–91, 2015.
- [45] D. Kliazovich, P. Bouvry, and S. U. Khan, "GreenCloud: A packet-level simulator of energy-aware cloud computing data centers," *J. Supercomput.*, vol. 62, no. 3, pp. 1263–1283, Dec. 2012.
- [46] S.-H. Lim, B. Sharma, G. Nam, E. K. Kim, and C. R. Das, "MDCSim: A multi-tier data center simulation, platform," in *Proc. IEEE Int. Conf. Cluster Comput. Workshops*, 2009, pp. 1–9.
- [47] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, "CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," *Softw., Pract. Exper.*, vol. 41, no. 1, pp. 23–50, Jan. 2011.
- [48] S. Bąk, M. Krystek, K. Kurowski, A. Oleksiak, W. Piątek, and J. Wąglarz, "GSSIM—A tool for distributed computing experiments," *Sci. Program.*, vol. 19, no. 4, pp. 231–251, 2011.
- [49] X. Jin, F. Zhang, A. V. Vasilakos, and Z. Liu, "Green data centers: A survey, perspectives, and future directions," 2016, *arXiv:1608.00687*. [Online]. Available: <http://arxiv.org/abs/1608.00687>
- [50] J. Shuja, K. Bilal, S. A. Madani, M. Othman, R. Ranjan, P. Balaji, and S. U. Khan, "Survey of techniques and architectures for designing energy-efficient data centers," *IEEE Syst. J.*, vol. 10, no. 2, pp. 507–519, Jun. 2016.
- [51] K. Bilal, S. U. Khan, S. A. Madani, K. Hayat, M. I. Khan, N. Min-Allah, J. Kolodziej, L. Wang, S. Zeadally, and D. Chen, "A survey on green communications using adaptive link rate," *Cluster Comput.*, vol. 16, no. 3, pp. 575–589, Jul. 2012.
- [52] K. Bilal, S. U. R. Malik, O. Khalid, A. Hameed, E. Alvarez, V. Wijaysekara, R. Irfan, S. Shrestha, D. Dwivedy, M. Ali, U. S. Khan, A. Abbas, N. Jalil, and S. U. Khan, "A taxonomy and survey on green data center networks," *Future Gener. Comput. Syst.*, vol. 36, pp. 189–208, Jul. 2014.
- [53] A. Hammadi and L. Mhamdi, "A survey on architectures and energy efficiency in data center networks," *Comput. Commun.*, vol. 40, pp. 1–21, Mar. 2014.
- [54] F. A. Moghaddam, P. Lago, and P. Grosso, "Energy-efficient networking solutions in cloud-based environments: A systematic literature review," *ACM Comput. Surveys*, vol. 47, no. 4, pp. 1–32, Jul. 2015.
- [55] F. Idzikowski, L. Chiaraviglio, A. Cianfrani, J. L. Vizcaino, M. Polverini, and Y. Ye, "A survey on energy-aware design and operation of core networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1453–1499, 2nd Quart., 2016.
- [56] F. Dabaghi, Z. Movahedi, and R. Langar, "A survey on green routing protocols using sleep-scheduling in wired networks," *J. Netw. Comput. Appl.*, vol. 77, pp. 106–122, Jan. 2017.
- [57] A. P. Bianzino, C. Chaudet, D. Rossi, and J.-L. Rougier, "A survey of green networking research," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 3–20, 1st Quart., 2012.
- [58] K. Cengiz and T. Dag, "A review on the recent energy-efficient approaches for the Internet protocol stack," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 1, pp. 1–22, Dec. 2015.
- [59] S.-Y. Jing, S. Ali, K. She, and Y. Zhong, "State-of-the-art research study for green cloud computing," *J. Supercomput.*, vol. 65, no. 1, pp. 445–468, Jul. 2013.
- [60] T. Mastelic, A. Oleksiak, H. Claussen, and I. Brandic, "Cloud computing: Survey on energy efficiency," *ACM Comput. Surveys*, vol. 47, no. 2, pp. 1–36, Dec. 2014.
- [61] K. Tejaswini and H. S. Guruprasad, "Energy efficiency in cloud computing: A review," *Int. J. Sci. Res. Manage. Stud.*, vol. 2, no. 5, pp. 253–259, 2015.
- [62] A. Majeed and M. A. Shah, "Energy efficiency in big data complex systems: A comprehensive survey of modern energy saving techniques," *Complex Adapt. Syst. Model.*, vol. 3, no. 1, p. 6, Dec. 2015.
- [63] F. Kong and X. Liu, "A survey on green-energy-aware power management for datacenters," *ACM Comput. Surv.*, vol. 47, no. 2, pp. 1–38, Jan. 2015.
- [64] S. P. Ahuja and K. Muthiah, "Survey of state-of-art in green cloud computing," *Int. J. Green Comput.*, vol. 7, no. 1, pp. 25–36, Jan. 2016.
- [65] J. Sekhar, G. Jeba, and S. Durga, "A survey on energy efficient server consolidation through VM live migration," *Int. J. Adv. Eng. Technol.*, vol. 5, no. 1, p. 515, Nov. 2012.
- [66] E. S. Mkoba and M. A. A. Saif, "A survey on energy efficient with task consolidation in the virtualized cloud computing environment," *Int. J. Res. Eng. Technol.*, vol. 3, no. 3, pp. 70–73, 2014.
- [67] E. C. Inacio and M. A. R. Dantas, "A survey into performance and energy efficiency in HPC, cloud and big data environments," *Int. J. Netw. Virtual Org.*, vol. 14, no. 4, pp. 299–318, 2014.
- [68] B. Pavithra and R. Ranjana, "A survey on energy aware resource allocation techniques in cloud," *Int. J. Sci., Eng. Technol. Res.*, vol. 4, no. 11, pp. 3751–3755, Nov. 2015.
- [69] G. Poonacha C, K. Priya K, P. Sheelvanthmath, and P. Ashwini, "A survey on energy aware scheduling of VMs in cloud data centers," *Int. J. Comput. Sci. Mobile Comput.*, vol. 4, no. 3, Mar. 2015.
- [70] S. Bose and J. Kumar, "A survey on energy aware load balancing techniques in cloud computing," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 5, pp. 397–402, May 2015.
- [71] P. Matre, D. S. Silakari, and U. Chourasia, "A survey on energy aware resource allocation for cloud computing," *Int. J. Eng. Res.*, vol. V5, no. 2, pp. 21–31, Feb. 2016.
- [72] M. A. Khoshkholghi, A. Abdullah, S. Subramaniam, M. Othman, and M. N. Derahman, "A taxonomy and survey of power management strategies in cloud data centers," *Int. J. Commun. Antenna Propag.*, vol. 6, no. 5, p. 305, 2016.
- [73] N. Shaik, "A survey on energy aware job scheduling algorithms in cloud environment," *I-Manager's J. Cloud Comput.*, vol. 3, no. 1, p. 30, Jan. 2016.
- [74] C. Jin, B. R. de Supinski, D. Abramson, H. Poxon, L. DeRose, M. N. Dinh, M. Endrei, and E. R. Jessup, "A survey on software methods to improve the energy efficiency of parallel computing," *Int. J. High Perform. Comput. Appl.*, vol. 31, no. 6, pp. 517–549, Sep. 2016.
- [75] S. F. Piraghaj, R. N. Calheiros, A. V. Dastjerdi, and R. Buyya, "A survey and taxonomy of energy efficient resource management techniques in platform as a service cloud," in *Handbook of Research on End-to-End Cloud Computing Architecture Design*. Harrisburg, PA, USA: IGI Global, 2017, pp. 410–454.
- [76] M. Splieth, F. Kramer, and K. Turowski, "Classification of techniques for energy efficient load distribution algorithms in clouds—a systematic literature review," in *Proc. 28th EnvironInfo Conf.*, Oldenburg, Germany, Sep. 2014, pp. 605–612.
- [77] X. Ma, Y. Cui, and I. Stojmeovic, "Energy efficiency on location based applications in mobile cloud computing: A survey," *Procedia Comput. Sci.*, vol. 10, pp. 577–584, Jan. 2012.

- [78] Y. Sharma, B. Javadi, W. Si, and D. Sun, "Reliability and energy efficiency in cloud computing systems: Survey and taxonomy," *J. Netw. Comput. Appl.*, vol. 74, pp. 66–85, Oct. 2016.
- [79] M. Karpowicz and E. Niewiadomska-Szynkiewicz, "Energy and power efficiency in cloud," in *Resource Management for Big Data Platforms*. Berlin, Germany: Springer, 2016, pp. 97–127.
- [80] M. Zakarya and L. Gillam, "Energy efficient computing, clusters, grids and clouds: A taxonomy and survey," *Sustain. Comput., Informat. Syst.*, vol. 14, pp. 13–33, Jun. 2017.
- [81] T. Xie, "SEA: A striping-based energy-aware strategy for data placement in RAID-structured storage systems," *IEEE Trans. Comput.*, vol. 57, no. 6, pp. 748–761, Jun. 2008.
- [82] R. T. Kaushik and M. Bhandarkar, "GreenHDFS: Towards an energy-conserving, storage-efficient, hybrid Hadoop compute cluster," in *Proc. Int. Conf. Power Aware Comput. Syst. (HotPower)*. Berkeley, CA, USA: USENIX Association, 2010, pp. 1–9.
- [83] R. T. Kaushik, L. Cherkasova, R. Campbell, and K. Nahrstedt, "Lightning: Self-adaptive, energy-conserving, multi-zoned, commodity green cloud storage system," in *Proc. 19th ACM Int. Symp. High Perform. Distrib. Comput. (HPDC)*. New York, NY, USA: ACM, 2010, p. 332.
- [84] X. D. You, C. Dong, L. Zhou, J. Huang, and C. F. Jiang, "Anticipation-based green data classification strategy in cloud storage system," *Appl. Math. Inf. Sci.*, vol. 9, no. 4, pp. 2151–2160, 2015.
- [85] T. Zhang, B. Liao, H. Sun, F. G. Li, and J. H. Ji, "Energy-efficient algorithm based on data classification for cloud storage system," *J. Comput. Appl.*, vol. 34, no. 8, pp. 2267–2273, 2014.
- [86] X. L. Xu, G. Yang, L. J. Li, and R. C. Wang, "Dynamic data aggregation algorithm for data centers of green cloud computing," *Syst. Eng. Electron.*, vol. 34, no. 9, pp. 1923–1929, Sep. 2012.
- [87] S. Q. Long, "Research on data layout strategies for cloud storage system," Ph.D. dissertation, School Comput., South China Univ. Technol., Guangzhou, China, 2014.
- [88] L. Bin, Y. Jiong, S. Hua, and N. Mei, "Energy-efficient algorithms for distributed storage system based on data storage structure reconfiguration," *J. Comput. Res. Develop.*, vol. 50, no. 1, pp. 3–18, 2013.
- [89] H. Amur, J. Cipar, V. Gupta, G. R. Ganger, M. A. Kozuch, and K. Schwan, "Robust and flexible power-proportional storage," in *Proc. 1st ACM Symp. Cloud Comput. (SoCC)*, Jun. 2010, pp. 217–228.
- [90] H. H. Le, S. Hikida, and H. Yokota, "An evaluation of power-proportional data placement for Hadoop distributed file systems," in *Proc. IEEE 9th Int. Conf. Dependable, Autonomic Secure Comput.*, Dec. 2011, pp. 752–759.
- [91] H. H. Le, S. Hikida, and H. Yokota, "Efficient gear-shifting for a power-proportional distributed data-placement method," in *Proc. IEEE Int. Conf. Big Data*, Oct. 2013, vol. E98.D, no. 5, pp. 76–84.
- [92] N. Maheshwari, R. Nanduri, and V. Varma, "Dynamic energy efficient data placement and cluster reconfiguration algorithm for MapReduce framework," *Future Gener. Comput. Syst.*, vol. 28, no. 1, pp. 119–127, Jan. 2012.
- [93] A. Wildani and E. L. Miller, "Semantic data placement for power management in archival storage," in *Proc. 5th Petascale Data Storage Workshop (PDSW)*, Nov. 2010, pp. 1–5.
- [94] R. Reddy, A. Kathpal, J. Basak, and R. Katz, "Data layout for power efficient archival storage systems," in *Proc. Workshop Power-Aware Comput. Syst. (HotPower)*, 2015, pp. 16–20.
- [95] X. Li, Y.-A. Tan, and Z. Sun, "Semi-RAID: A reliable energy-aware raid data layout for sequential data access," in *Proc. 27th Symp. Massive Storage Technol. (MSST)*, 2011, pp. 1–11.
- [96] Y. W. Xiao, J. B. Wang, Y. P. Li, and H. Gao, "Energy efficient data placement algorithm and node-scheduling policies in cloud computing systems," *J. Comput. Res. Develop.*, vol. 50, pp. 342–351, Jul. 2013.
- [97] R. Xiong, J. Luo, and F. Dong, "SLDP: A novel data placement strategy for large-scale heterogeneous Hadoop cluster," in *Proc. 2nd Int. Conf. Adv. Cloud Big Data*, Nov. 2014, pp. 9–17.
- [98] R. Xiong, J. Luo, and F. Dong, "Optimizing data placement in heterogeneous Hadoop clusters," *Cluster Comput.*, vol. 18, no. 4, pp. 1465–1480, Dec. 2015.
- [99] J. Song, Z. Wang, T. T. Li, and G. Yu, "Energy consumption optimization data placement algorithm for MapReduce system," *J. Softw.*, vol. 26, no. 8, pp. 2091–2110, 2015.
- [100] F. Lemma and C. Fetzer, "dyn-PowerCass: Energy efficient distributed store based on dynamic data placement strategy," in *Proc. AFRICON*, Sep. 2015, pp. 1–5.
- [101] W. Lang, J. M. Patel, and J. F. Naughton, "On energy management, load balancing and replication," *ACM SIGMOD Rec.*, vol. 38, no. 4, pp. 35–42, Dec. 2009.
- [102] J. Kim and D. Rotem, "Using replication for energy conservation in RAID systems," in *Proc. Int. Conf. Parallel Distrib. Process. Techn. Appl.*, 2010, pp. 703–709.
- [103] E. Thereska, A. Donnelly, and D. Narayanan, "Sierra: Practical power-proportionality for data center storage," in *Proc. 6th Conf. Comput. Syst. (EuroSys)*, Apr. 2011, pp. 169–182.
- [104] Y. Y. Liu, "Research on energy efficient replication management and task scheduling under cloud environment," M.S. thesis, School Comput. Sci. Technol., Nanjing Univ. Aeronaut. Astronaut., Nanjing, China, 2011.
- [105] L. G. Zhang and Y. H. Deng, "Designing a power-aware replication strategy for storage clusters," in *Proc. IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Social Comput.*, Aug. 2013, pp. 224–231.
- [106] X. Y. Luo, "Research on replica placement algorithms towards high-efficient clustered storage systems," Ph.D. dissertation, School Comput. Sci., Southeast Univ., Dhaka, Bangladesh, 2013.
- [107] S. Long, Y. Zhao, and W. Chen, "A three-phase energy-saving strategy for cloud storage systems," *J. Syst. Softw.*, vol. 87, pp. 38–47, Jan. 2014.
- [108] S.-Q. Long, Y.-L. Zhao, and W. Chen, "MORM: A multi-objective optimized replication management strategy for cloud storage cluster," *J. Syst. Archit.*, vol. 60, no. 2, pp. 234–244, Feb. 2014.
- [109] X. L. Cui, B. Mills, T. Znati, and R. Melhem, "Shadow replication: An energy-aware, fault-tolerant computational model for green cloud computing," *Energies*, vol. 7, pp. 5151–5176, Aug. 2014.
- [110] Z. Y. Wang, J. Yu, C. T. Ying, L. Lu, and A. Q. Ban, "Energy-efficient strategy for dynamic management of cloud storage replica based on user visiting characteristic," *J. Comput. Appl.*, vol. 34, 8, pp. 2256–2259 and 2290, 2014.
- [111] D. Boru, D. Kliazovich, F. Granelli, P. Bouvry, and A. Y. Zomaya, "Energy-efficient data replication in cloud computing datacenters," *Cluster Comput.*, vol. 18, no. 1, pp. 385–402, Mar. 2015.
- [112] Y. H. Lin and H. Y. Shen, "EAfr: An energy-efficient adaptive file replication system in data-intensive clusters," in *Proc. 24th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2015, pp. 1–8.
- [113] Z. L. Shi, "Research on efficient replica management strategy in cloud environment," M.S. thesis, School Comput., Univ. Electron. Sci. Technol. China, Chengdu, China, 2015.
- [114] D. Kliazovich, P. Bouvry, and F. Granelli, *Energy Consumption Optimization in Cloud Data Centers*. Piscataway, NJ, USA: IEEE, 2015.
- [115] B. A. Milani and N. J. Navimipour, "A comprehensive review of the data replication techniques in the cloud environments: Major trends and future directions," *J. Netw. Comput. Appl.*, vol. 64, pp. 229–238, Apr. 2016.
- [116] D. Sun and R. Huang, "A stable online scheduling strategy for real-time stream computing over fluctuating big data streams," *IEEE Access*, vol. 4, pp. 8593–8607, 2016.
- [117] X. You, Y. Li, M. Zheng, C. Zhu, and L. Yu, "A survey and taxonomy of energy efficiency relevant surveys in cloud-related environments," *IEEE Access*, vol. 5, pp. 14066–14078, 2017.
- [118] R. Yadav, W. Zhang, O. Kaiwartya, P. R. Singh, I. A. Elgendy, and Y.-C. Tian, "Adaptive energy-aware algorithms for minimizing energy consumption and SLA violation in cloud computing," *IEEE Access*, vol. 6, pp. 55923–55936, 2018.
- [119] L. Wang, Y. Ma, J. Yan, V. Chang, and A. Y. Zomaya, "PipsCloud: High performance cloud computing for remote sensing big data management and processing," *Future Gener. Comput. Syst.*, vol. 78, pp. 353–368, Jan. 2018.
- [120] J. Song, H. He, Z. Wang, G. Yu, and J.-M. Pierson, "Modulo based data placement algorithm for energy consumption optimization of MapReduce system," *J. Grid Comput.*, vol. 16, no. 3, pp. 409–424, Sep. 2018.
- [121] X. T. Tran, T. V. Do, C. Rotter, and D. Hwang, "A new data layout scheme for energy-efficient MapReduce processing tasks," *J. Grid Comput.*, vol. 16, no. 2, pp. 285–298, Jun. 2018.
- [122] Y. Ebadati and N. J. Navimipour, *An Energy-Aware Method for Data Replication in the Cloud Environments Using a Tabu Search and Particle Swarm Optimization Algorithm*. Hoboken, NJ, USA: Wiley, 2018.
- [123] J. Yu, Z. Kuang, B. Zhang, W. Zhang, D. Lin, and J. Fan, "Leveraging content sensitiveness and user trustworthiness to recommend fine-grained privacy settings for social image sharing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1317–1332, May 2018.

[124] G. Jia, G. Han, J. Du, and S. Chan, "A maximum cache value policy in hybrid memory-based edge computing for mobile devices," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4401–4410, Jun. 2019.

[125] R. Vales, J. Moura, and R. Marinheiro, "Energy-aware and adaptive fog storage mechanism with data replication ruled by spatio-temporal content popularity," *J. Netw. Comput. Appl.*, vol. 135, pp. 84–96, Jun. 2019.

[126] H. Gao, Y. Xu, Y. Yin, W. Zhang, R. Li, and X. Wang, "Context-aware QoS prediction with neural collaborative filtering for Internet-of-Things services," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4532–4542, May 2020.

[127] Y. Zhu, W. Zhang, Y. Chen, and H. Gao, "A novel approach to workload prediction using attention-based LSTM encoder-decoder network in cloud environment," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 274, Dec. 2019.

[128] Y. Yin, J. Xia, Y. Li, Y. Xu, W. Xu, and L. Yu, "Group-wise itinerary planning in temporary mobile social network," *IEEE Access*, vol. 7, pp. 83682–83693, 2019.

[129] G. Jia, G. Han, J. J. P. C. Rodrigues, J. Lloret, and W. Li, "Coordinate memory deduplication and partition for improving performance in cloud computing," *IEEE Trans. Cloud Comput.*, vol. 7, no. 2, pp. 357–368, Apr. 2019.

[130] Y. Yin, L. Chen, Y. Xu, J. Wan, H. Zhang, and Z. Mai, "QoS prediction for service recommendation with deep feature learning in edge computing environment," *Mobile Netw. Appl.*, vol. 25, no. 2, pp. 391–401, Apr. 2020.

[131] G. Jia, G. Han, H. Xie, and J. Du, "Hybrid-LRU caching for optimizing data storage and retrieval in edge computing-based wearable sensors," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1342–1351, Apr. 2019.

[132] L. Kuang, T. Gong, S. OuYang, H. Gao, and S. Deng, "Offloading decision methods for multiple users with structured tasks in edge computing for smart cities," *Future Gener. Comput. Syst.*, vol. 105, pp. 717–729, Apr. 2020.

[133] H. Gao, C. Liu, Y. Li, and X. Yang, "V2 VR: Reliable hybrid-network-oriented V2 V data transmission and routing considering RSUs and connectivity probability," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–14, 2020.

[134] G. Jia, G. Han, J. Jiang, L. Liu, and L. Shu, "DPAM: A demand-based page-level address mappings algorithm in flash memory for smart industrial edge devices," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1993–2002, Mar. 2020.



XUEQIANG LV received the Ph.D. degree from Northeastern University, in 2003. He held a post-doctoral position with Peking University, from 2003 to 2005. He is currently a Professor with the Beijing Information Science & Technology University. He has been in charge of the National Nature Science Foundation of China three times. He has authored about 60 articles in the international conference or journals, most of them are indexed by EI or SCI database. His current research areas include cloud computing, distributed computing, natural language processing, image processing, information retrieval, machine learning, and deep learning.



ZHIKAI ZHAO received the Ph.D. degree from the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, in 2012. He is currently a full-time Researcher with the Internet of Things (IoT) Research Center, and the National Joint Engineering Laboratory of Internet Applied Technology of Mines, China University of Mining and Technology. He has published more than 20 articles in related international conferences and journals. His main research interests include the IoT, machine learning, pattern recognition, and intelligent information processing.



JUNMEI HAN received the M.S. degree in military communication and the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2013 and 2019, respectively. She is currently an Associate Professor with the Department of Systems General Design Institute of Systems Engineering, AMS, PLA, China. Her current research areas include complex system theory, natural language processing, and graph neural network.



XUEPING REN received the B.S. degree from Hanzhou Dianzi University, Hangzhou, China, in 2005. She is currently an Associate Professor with the Department of Computer Science, Hanzhou Dianzi University, China. She has been an In Charge of the Nature Science Funding of Zhejiang Province, since 2018. She has authored more than ten articles in the international conference or journals, most of them are indexed by EI or SCI database. Her current research areas include privacy and security of RFID, WSN, and energy management.

...



XINDONG YOU received the Ph.D. degree with Northeastern University, in 2007. She was an Associate Professor with Hangzhou Dianzi University. She held a postdoctoral position with Tsinghua University with the Beijing Institute of Graphic Communication, from 2016 to 2018. She is currently an Associate Professor with the Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science & Technology University, China. She is in charge of the National Nature Science Foundation of China, from 2014 to 2017, and the Nature Science Funding of Zhejiang Province, from 2013 to 2015. She has authored about 30 articles in the international conference or journals, most of them are indexed by EI or SCI database. Her current research areas include natural language processing, image processing, distributed computing, cloud storage, energy management, and data replica management.