Author/s:
You, X;Sun, T;Sun, D;Liu, X;Lv, X;Buyya, R

Title:
K-ear: Extracting data access periodic characteristics for energy-aware data clustering and storing in cloud storage systems

Date:
2021-05-10

Citation:
You, X., Sun, T., Sun, D., Liu, X., Lv, X. & Buyya, R. (2021). K-ear: Extracting data access periodic characteristics for energy-aware data clustering and storing in cloud storage systems. Concurrency and Computation Practice and Experience, 33 (9), https://doi.org/10.1002/cpe.6096.

You Xindong (Orcid ID: 0000-0002-3351-4599)
Sun Dawei (Orcid ID: 0000-0003-3137-6257)
Buyya Rajkumar (Orcid ID: 0000-0001-9754-6496)

K-ear: Extracting Data Access Periodic Characteristics for Energy-aware Data Clustering and Storing in Cloud Storage Systems

Xindong You [1], Tian Sun[1], Dawei Sun[2], Xunyun Liu[3], Xueqiang Lv[1], and Rajkumar Buyya[4]

1. *Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science and Technology University, Beijing, China*
2. *School of Information Engineering, China University of Geosciences, Beijing, China*
3. *Artificial Intelligence Research Center, National Innovation Institute of Defense Technology, Beijing, China*
4. *Cloud Computing and Distributed Systems (CLOUDS) Lab, School of Computing and Information Systems, The University of Melbourne, Australia*

* Corresponding author: Dawei Sun

Abstract

Rapid increase in energy consumption is a serious problem in cloud storage systems. Data accessed in large-scale storage systems usually exhibit temporal and spatial characteristics, which make it possible to reduce energy consumption by clustering data with similar access characteristics for storage in the same zone of cloud storage systems. Existing works usually only focus on the frequency of data access. However, widely existing phenomena show data access with seasonal and tidal characteristics in cloud storage systems. The seasonal and tidal characteristics of data access are extracted thoroughly in this paper. According to the extracted data access characteristics, energy-aware data clustering through a machine learning algorithm (K-ear) is proposed. K-ear classifies data into five seasonal categories according to their seasonal access characteristics and then classifies every seasonal category into three tidal categories according to its tidal access characteristics. The fifteen classified categories are stored in different storage zones with different energy and performance modes. Simulation experiments using CloudSimDisk with the constructed mathematical models demonstrate that the proposed K-ear algorithm is more energy-efficient than the default data clustering algorithms in Hadoop and the classical data clustering storage strategy according to the data access frequency (SEA, Striping-Based Energy-Aware Strategy)

Keywords:

Energy Consumption, Cloud Storage System, Data Clustering, Data Access Characteristics, Seasonal Characteristics, Tidal Characteristics.

1. Introduction

The exponential growth of the volume data is becoming one of the leading causes of high energy consumption in cloud storage systems [1][2][3]. It has been reported in the literature [4] that the energy consumed by data centers will be more than 1000 TWh during 2013-2025, which will surpass the total energy consumption of Japan and Germany. The energy consumed by data centers, including for cooling equipment, will consume 5% of the total energy consumption worldwide. An even more serious problem is that increasing energy consumption will produce high carbon and GHG (Greenhouse Gases) emissions [6][7], which will result in serious environmental pollution. Therefore, reducing energy consumption is one of the hottest issues in the cloud storage domain. Classifying data into different categories according to their access characteristics is an efficient way to enhance energy efficiency in cloud storage systems. Different data categories are stored in different storage zones, running in different energy and performance modes. This kind of method will lead to less energy consumption in the storage zone, which has a low energy and performance mode while the workload is light. However, only the instant access frequency is considered in the current existing energy-aware data classification strategies. Long periodic access characteristics, such as seasonal characteristics and tidal characteristics, are not considered. Classifying data according to their instant access frequency will lead to frequent data migration, which will result in worse performance. Extracting the long periodic access characteristics is an effective solution, which could allow the low energy and performance running mode for a relatively long time. On the other hand, seasonal and tidal access characteristics obviously appear in cloud storage systems. As shown in Fig.1, the access frequencies (search index) of spring clothing, summer clothing, autumn clothing and winter clothing exhibit obvious seasonal period characteristics. Other words with seasonal characteristics also have seasonal periodic access frequency.

As Fig 1 is the screenshot from the Baidu website, there are Chinese words in the figure. We list the Chinese words in the Fig.1 from left to right, and top to down. where "春装"means "spring clothings", "夏装"-"summer clothings", "秋装" –"autumn clothings", "冬装" winter clothings. "平均值" –"average value", "搜索指数"-"search indexes".

As Fig 2 is also the screenshot from the Baidu website, there are Chinese words in the figure. We list the Chinese words in the Fig.2 from left to right, and top to down. where "整体趋势"means "whole trend", "PC 趋势"-"PC trend", "移动趋势" –"mobile trend", "最近"-"recent", "7 天"-"7 days", "30 天"-"30days", "90 天"-"90 days", "半年"-"half a year", "全部"-"all", "自定义"-"custom","平均值" –"average value", "搜索指数"-"search indexes".

As shown in Fig. 2, the access frequency of work-related words has the same tidal periodic access characteristics: the peak points usually appear on workdays and the valley points almost always appear on weekends. This kind of weekly tidal period phenomenon always appears in work-related words, documents and media.

As Fig3 is also the screenshot from the Baidu website, there are Chinese words in the figure. We list the Chinese words in the Fig.s from left to right, and top to down. where "整体趋势"means "whole trend", "PC 趋势"-"PC trend", "移动趋势" –"mobile trend", "最近"-"recent", "7 天"-"7 days", "30 天"-"30days", "90 天"-"90 days", "半年"-"half a year", "全部"-"all", "自定义"-"custom", "电视剧"-"TV", "电影"-"movie", "游戏"-"game","平均值" –"average value", "搜索指数"-"search indexes".

On the other hand, as shown in Fig. 3, the access frequency of entertainment-related words also has the same tidal periodic access characteristics. The peak points usually appear on weekends, and the valley points almost always appear on workdays. This kind of weekly tidal period phenomenon always appears in entertainment-related words, documents and media.

Recently, placing data according to the access characteristics is one of the main techniques for energy saving in large scale storage systems [8]~[11]. However, only the temporal data access characteristics are considered in the existing literatures.

Based on the above observations and due to a lack of consideration of the periodic data access characteristics of the current existing energy-aware data classification work, we focused on extracting the seasonal and tidal access characteristics for energy-aware data clustering storage in this paper. In this direction, our paper makes the following key contributions:

(1) Data seasonal and tidal access characteristics extracting algorithms (SCEA and TCEA) are designed, in which how to store the data access frequency and how to express the seasonal and tidal characteristics are described in detail.
(2) The energy-aware data clustering strategy K-ear is designed, in which the framework of K-ear is first constructed and then the data are clustered by an unsupervised machine learning algorithm into different categories. Correspondingly, the cloud storage system is divided into fifteen zones to store data with similar access characteristics.

(3) The proposed K-ear algorithm is modeled, and the classical data classification SEA and the Hadoop default data placement algorithms are used for comparison through a mathematical method. The constructed mathematics model is the basis for analyzing the energy efficiency of the proposed K-ear strategy and is also used to conduct simulation experiments to verify the advantage of the proposed strategy in energy consumption.

(4) Substantial simulation experiments are conducted using extended CloudSimDisk simulator, and the results demonstrate that the proposed K-ear strategy is more energy-efficient than the other two data placement algorithms.

The rest of the paper is organized as follows. In Section 2, we analyze the related energy-aware data classification work. The detailed framework and the algorithms of K-ear are described in Section 3. The mathematical model of the proposed K-ear strategy and the classical energy-aware data classification algorithm SEA are constructed for energy and performance analysis in Section 4. In Section 5, we demonstrate the detailed energy consumption and performance of the proposed K-ear, SEA and Hadoop default data storage algorithms. Conclusions and future work are presented in the final section.

2. Related Work

Energy-aware data classification strategies classify data into different categories according to the data access characteristics and then partition the storage system into different zones. Data with similar access characteristic clustering are stored in the same storage zone. Different zones run in different energy and performance modes. Energy consumption savings are achieved by managing the power state of the different zones. T. Xie first proposed a striping-based energy-aware data placement strategy (SEA) in the RAID (Redundant Arrays of Independent Disks) storage system [12], in which RAID disks are divided into Hot Disk Zone and Cold Disk Zone. Popular data are stored in the Hot Disk Zone while unpopular data are stored in the Cold Disk Zone. Disks in Hot Disk Zone run in a mode with a high transfer rate and high power rate, while disks in Cold Disk Zone run in a mode with a low transfer rate and low power rate. Analysis and simulation results show that the proposed SEA mechanism saves much energy consumption with little performance loss. Analyzing the access traces in Yahoo showed that the data access patterns in the Hadoop cluster have obvious heterogeneity. R.T. Kaushik et al. designed the GreenHDFS mechanism, in which data are classified into different categories according to temperature. The temperature of data is deduced by their availability and the user's performance requirement. Correspondingly, the Hadoop cluster is divided into Hot Zone and Cold Zone. Simulation experiments conducted on the generated workload based on a real trace of three months of data from Yahoo demonstrate that 26% of energy consumption can be reduced by only managing the cold zone at a lower power consumption rate while the system load is light [13]. A similar energy-aware data classification policy, Lightning, was designed by the Kaushik team [14] to reduce the energy consumption of the Yahoo cloud storage system. Inspired by GreenHDFS and the lighting mechanism, we have proposed a green data classification strategy based on anticipation named AGDC, in which the neural network is employed to predict the temperature of data. Based on the predicted temperature of data, they are classified into three categories: cold data, seasonal hot data and hot data. Correspondingly, the cloud storage system is divided into different zones. Simulation experiments conducted on the GridSim simulator demonstrated that the AGDC mechanism can save approximately 16% of energy consumption at the expense of an increased average response time of 0.005 s. AGDC outperformed the integration general classification algorithm TDCS [15]. In the literature [19], data are classified into different categories according to their access frequency and regularity. RACK is divided into Active-Zone and Sleep-Zone. Data with different access characteristics are stored in different zones. Simulation experimental results in the MATLAB and GridMix environments show that the energy consumption saved by the proposed algorithm is up to 39.01%. An energy-efficient algorithm that classifies data in the cloud storage system was proposed by Z. Tao et al. [16] that divided the cloud storage system into HotZone, ColdZone and Reduplication Zone. Data are stored in the zones according to their repetition and activity factor characteristics. The experimental results show that the energy utilization rate improved by 25%. Furthermore, the proposed algorithm performs better when the system load is light. Recently, the SLA (Service Level Agreement) has been considered in more and more literatures for trading off the energy efficiency and the QoS. A dynamic data aggregation algorithm for green cloud computing is proposed in [17]. According to the data access pattern, the data are aggregately stored dynamically among nodes. By managing the power state of the storage nodes, energy consumption can be reduced along with the QoS considered. Dr. Long designed static and dynamic file layout and replica and data layout policies to reduce the energy consumption in cloud storage systems [18]. The static file layout strategy (SFLS) first divides data into hot files and big files according to their access frequency and service time. Correspondingly, the disks are divided into different groups. I/O requests are distributed to the different disk groups according to the access frequency and service time. The results obtained from the CloudSim simulator demonstrate that SFLS can save power consumption by over 35% compared to the default HDFS. Moreover, R.Yadav et la published the related article for minimizing energy consumption and SLA violation in cloud computing[23]~[26]. Other data placement or data layout strategies for energy efficiency have been published in recent years [19]~[35], but the period access characteristics also have not been extracted for data clustering storing.

As described before, data classification is an efficient way to reduce energy consumption. However, only the instant data access frequency is considered in all of the above energy-aware data classification strategies, which may cause frequent data migration and result in performance loss. The seasonal and tidal periodic access characteristics of data are thoroughly extracted for data classification storing in our proposed K-ear strategy, which can leave the nodes in a low power state for a relatively long time, leading to energy consumption savings.

A summary of the reviewed related work is presented in Table 1, comparing in terms of whether the access frequency or the periodic access characteristics (seasonal or tidal periodic access characteristics) has been considered or not, the zones have been divided.

3. K-ear: Energy-aware Data Clustering Strategy

The proposed K-ear strategy consists of two data access characteristics extraction algorithms (Seasonal Characteristics Extracting Algorithm SCEA, and Tidal Characteristics Extracting Algorithm TCEA) and a framework, which are described in the following subsections.

3.1 Periodic Data Access Characteristics Extraction

The data seasonal and tidal access characteristics are extracted by the SCEA and TCEA algorithms, respectively. To explain the SCEA and TCEA algorithms clearly, some definitions are given first.

Representation Data Set: $\mathrm{D}=\{\mathrm{d}_1,\mathrm{d}_2,\cdots,\mathrm{d}_m\}$, which is the representation data set of the primary data for clustering, and m is the number of data.

Data Seasonal Characteristics: are represented by the vector $\mathrm{SE}=\begin{bmatrix}\mathrm{Se}_1\\ \mathrm{Se}_2\\ \vdots\\ \mathrm{Se}_m\end{bmatrix}$. Assume that y is the number of years of data to be collected and that there are four seasonal search index ratios for each year. Therefore, $\mathrm{Se}_i=[se_1,se_2,\cdots,se_{4*y}]$.

Data Tidal Characteristics: are represented by the vector $\mathrm{CX}=\begin{bmatrix}cx_1\\ cx_2\\ \vdots\\ cx_m\end{bmatrix}$ and $c\mathrm{x}_i=[\mathrm{p}_{\mathrm{i},1}\quad \mathrm{v}_{\mathrm{i},1}\quad \mathrm{p}_{\mathrm{i},2}\quad \mathrm{v}_{\mathrm{i},2}\quad \cdots\quad \mathrm{p}_{\mathrm{i,z}}\quad \mathrm{v}_{\mathrm{i,z}}]$, in which z is the number of weeks of representation data (there are 52 weeks in one year).

The seasonal characteristics extracting algorithm is described in Algorithm 1, through which the seasonal characteristics of the data stored in the SE vector.

Tidal characteristics extracting algorithm is described in Algorithm 2, through which the seasonal characteristics of the data are stored in the CX vector..

3.2 Framework and Strategy of K-ear

The main procedures of the proposed K-ear are delineated in Fig. 4. The seasonal access characteristics of data are extracted by the SCEA algorithm (see Algorithm 1 for details). Then, the machine learning clustering-related algorithm is employed on the extracted seasonal access characteristics. Data are clustered into five categories: spring data, summer data, autumn data, winter data and other data. Correspondingly, the cloud storage system is divided into five zones: the spring zone, summer zone, autumn zone, winter zone and other zones. The different data categories are stored in the corresponding storage zones. When entering a certain zone, the TCEA (Algorithm 2 for detail) is utilized to extract the tidal access characteristics of data. Similarly, the machine learning clustering-related algorithm is employed on the extracted tidal access characteristics. The data are further clustered into three categories: work-related data, entertainment data and other data. Therefore, the cloud storage system is partitioned into fifteen zones: the spring-work zone, spring-entertainment zone, spring-other zone, summer-work zone, summer-entertainment zone, summer-other zone, autumn-work zone, autumn-entertainment zone, autumn-other zone, winter-work zone, winter-entertainment zone, winter-other zone, other-work zone, other-entertainment zone, and other-other zone. We assume that a disk has two speed modes: a high-speed mode with a high transfer rate and a high energy-consuming rate and a low-speed mode with a low transfer rate and a low energy consumption rate, as it is a common method for modeling hard-disk power consumption [12][38][39][40]. According to the different periods of the year, different storage zones run in different speed modes. For example, on a workday in the spring, the spring-work zone, other-work zone and other-other zone run in the high speed mode for performance consideration. The other zones run in low-speed mode to save energy consumption.

The symbols used in the proposed K-ear algorithm are explained in Table 2 and Table3

When the s in the notation η_s is replaced respectively by m, a, and w, notations meaning the related characteristics of the Summer, Autumn and Winter respectively.

And the above symbols have the following relationships: $\eta_s + \eta_m + \eta_a + \eta_w + \eta_o = 1$,$\eta_{s-\mathrm{w}} + \eta_{s-\mathrm{h}} + \eta_{s-\mathrm{o}} = 1$,$\eta_{m-\mathrm{w}} + \eta_{m-\mathrm{h}} + \eta_{m-\mathrm{o}} = 1$,$\eta_{a-\mathrm{w}} + \eta_{a-\mathrm{h}} + \eta_{a-\mathrm{o}} = 1$, $\eta_{w-\mathrm{w}} + \eta_{w-\mathrm{h}} + \eta_{w-\mathrm{o}} = 1$,$\eta_{o-\mathrm{w}} + \eta_{o-\mathrm{h}} + \eta_{o-\mathrm{o}} = 1$.

Assume the number of the disks is n, and the disks set is defined as below. $\mathrm{K} = \{\mathrm{s}_1, \mathrm{s}_2, \cdots, \mathrm{s}_{n1}, \mathrm{m}_1, \mathrm{m}_2, \cdots \mathrm{m}_{n2}, \mathrm{a}_1, \mathrm{a}_2, \cdots \mathrm{a}_{n3}, w_1, w_2, \cdots, w_{n4}, \mathrm{o}_1, \mathrm{o}_2, \cdots \mathrm{o}_{n5}\}$

And the set $S = \{\mathrm{s}_1, \mathrm{s}_2, \cdots, \mathrm{s}_{n1}\}$ $M = \{\mathrm{m}_1, \mathrm{m}_2, \cdots \mathrm{m}_{n2}\}$ $A = \{\mathrm{a}_1, \mathrm{a}_2, \cdots \mathrm{a}_{n3}\}$ $W = \{w_1, w_2, \cdots, w_{n4}\}$ $O = \{\mathrm{o}_1, \mathrm{o}_2, \cdots \mathrm{o}_{n5}\}$ represent the disks store the data with spring, summer, autumn, winter and no seasonal characteristics respectively, in which $\mathrm{n}_1 = \eta_s \times n$,$\mathrm{n}_2 = \eta_m \times n$,$\mathrm{n}_3 = \eta_a \times n$,$\mathrm{n}_4 = \eta_w \times n$,$\mathrm{n}_5 = \eta_o \times n$. And the number of the disks to store the data with tidal characteristics of every seasonal zone are listed in Table 2.

In Table 3, when the s in the notation η_s is respectively replaced by m, a, and w, && n_1 is respectively replcaed by n_2, n_3, $n_4$notations meaning the number of the disk to store the data with related characteristics of the Summer, Autumn and Winter. Respectively.

The detailed procedures of the K-ear strategy are illustrated in Algorithm 3.

According to the algorithm 3, the number of disks running in different modes during different time zones are listed in Table 4.

4. Mathematic Modeling

To analyze the energy efficiency advantage of our proposed K-ear strategy, we model the K-ear mathematically in this section. The parameters used during mathematical modeling are listed in following tables. Parameters of the whole system are listed in Table5. Parameters about the spring season are listed in Table 6. Parameters about the summer season, autumn season and winter season will be explained correspondingly.

For short, when the *spring* in the labels of the Table 6 is replaced by the *summer*, *autumn* and *winter* respectively, the parameters are about the Summer season, Autumn season and Winter season respectively. Such as $spring\_work\_t_h^{active}$ means the total time of disks running in high mode with active status during workdays in Spring, and the $summer\_work\_t_h^{active}$ means the total time of disks running in high mode with active status during workdays in Summer. Furthermore, when the *S* in the labels of the Table 6 is replaced by the *M*, *A*, *W* respectively, the parameters are about the Summer season, Autumn season and Winter season respectively. Such as $S\_W_h^n$ means the total access times in high mode during workdays in Spring, and the $M\_W_h^n$ means the total access times in high mode during workdays in Summer. The replacement rule is the same for the remaining parameters.

According to the proposed K-ear strategy in Algorithm 3, the energy consumption model of K-ear is deduced as below.

$$\mathrm{e}_{total} = spring\_work\_\mathrm{e}_{total} + spring\_holiday\_\mathrm{e}_{total} + summer\_work\_\mathrm{e}_{total} + summer\_holiday\_\mathrm{e}_{total} + autumn\_work\_\mathrm{e}_{total} + autumn\_holiday\_\mathrm{e}_{total} + winter\_work\_\mathrm{e}_{total} + winter\_holiday\_\mathrm{e}_{total} \tag{1}$$

The intuition behind the equations (1) is that the total energy consumption of the systems is the summary of the energy consumption of the workdays in spring ($spring\_work\_\mathrm{e}_{total}$), holidays in spring ($spring\_holiday\_\mathrm{e}_{total}$), workdays in summer ($summer\_work\_\mathrm{e}_{total}$), holidays in summer ($summer\_holiday\_\mathrm{e}_{total}$), workdays in autumn ($autumn\_work\_\mathrm{e}_{total}$), holidays in autumn ($autumn\_holiday\_\mathrm{e}_{total}$), workdays in winter ($winter\_work\_\mathrm{e}_{total}$), holidays in winter ($winter\_holiday\_\mathrm{e}_{total}$). And there are calculated by the following formulas

$$spring_{work_{\mathrm{e}_{total}}} = spring\_work\_e_h^{active} + spring\_work\_e_h^{\mathrm{idle}} + spring\ \_work\_e_l^{active} + spring\_work\_e_l^{idle} \tag{2}$$

$$summer\_work\_\mathrm{e}_{total} = summer\_work\_e_h^{active} + summer\_work\_e_h^{idle} + summer\_work\_e_l^{active} + summer\_work\_e_l^{idle} \tag{3}$$

$$autumn\_work\_\mathrm{e}_{total} = autumn\_work\_e_h^{active} + autumn\_work\_e_h^{idle} + autumn\_work\_e_l^{active} + autumn\_work\_e_l^{idle} \tag{4}$$

$$winter\_work\_\mathrm{e}_{total} = winter\_work\_e_h^{active} + winter\_work\_e_h^{idle} + winter\_work\_e_l^{active} + winter\_work\_e_l^{idle} \tag{5}$$

$$spring\_holiday\_\mathrm{e}_{total} = spring\_holiday\_e_h^{active} + spring\_holiday\_e_h^{\mathrm{idle}} + spring\ \_holiday\_e_l^{active} + spring\_holiday\_e_l^{idle} \tag{6}$$

$$summer\_holiday\_\mathrm{e}_{total} = summer\_holiday\_e_h^{active} + summer\_holiday\_e_h^{idle} + summer\_holiday\_e_l^{active} + summer\_holiday\_e_l^{idle} \tag{7}$$

$$autumn\_holiday\_\mathrm{e}_{total} = autumn\_holiday\_e_h^{active} + autumn\_holiday\_e_h^{idle} + autumn\_holiday\_e_l^{active} + autumn\_holiday\_e_l^{idle} \tag{8}$$

$$winter\_holiday\_\mathrm{e}_{total} = winter\_holiday\_e_h^{active} + winter\_holiday\_e_h^{idle} + winter\_holiday\_e_l^{active} + winter\_holiday\_e_l^{idle} \tag{9}$$

When

$$\boldsymbol{spring\_work\_e_h^{active}} = \mathrm{p}^h \times spring\_work\_t_h^{active} = \mathrm{p}^h \times S\_W_h^n \times \mathrm{s}'/(\mathrm{\tau}^h \times H1) \tag{10}$$

$$\boldsymbol{spring\_work\_e_h^{idle}} = i^h \times (spring\_work\_T \times H1 - spring\_work\_t_h^{active}) = i^h \times (\frac{5T}{28} \times H1 - S\_W_h^n \times s'/(\tau^h \times H1) \quad (11)$$

$$\boldsymbol{spring\_holiday\_e_h^{active}} = p^h \times spring\_holiday\_t_h^{active} = p^h \times S\_H_h^n \times s'/(\tau^h \times H2) \quad (12)$$

$$\boldsymbol{spring\_holiday\_e_h^{idle}} = i^h \times (spring\_holiday\_T \times H2 - spring\_holiday\_t_h^{active}) = i^h \times (\frac{2T}{28} \times H2 - S_{H_h}{}^n \times s'/(\tau^h \times H2)) \quad (13)$$

And the other formulas to calculate the valuses related to the Summer season, Autumn season and Winter season are simluar to the formulas from (10) to (13), as they accorrding to the above replacement rules plus that H1 and H2 are replcaed by H3 and H4, H5 and H6, H7 and H8.

$$\boldsymbol{spring\_work\_e_l^{active}} = p^l \times spring\_work\_t_l^{active} = p^l \times S\_W_l^n \times s'/(\tau^l \times L1) \quad (14)$$

$$\boldsymbol{spring\_work\_e_l^{idle}} = i^l \times (spring\_work\_T \times L1 - spring\_work\_t_l^{active}) = i^l \times (\frac{5T}{28} \times L1 - S\_W_l^n \times s'/(\tau^l \times L1)) \quad (15)$$

$$\boldsymbol{spring\_holiday\_e_l^{active}} = p^l \times spring\_holiday\_t_l^{active} = p^l \times S\_H_l^n \times s'/(\tau^l \times L2) \quad (16)$$

$$\boldsymbol{spring\_holiday\_e_l^{idle}} = i^l \times (spring\_holiday\_T \times L2 - spring\_holiday\_t_l^{active}) = i^l \times (\frac{2T}{28} \times L2 - S\_H_l^n \times s'/(\tau^l \times L2)) \quad (17)$$

And the other formulas to calculate the valuses related to the Summer season, Autumn season and Winter season are simluar to the formulas from (14) to (17), as they accorrding to the above replacement rules plus that L1 and L2 are replcaed by L3 and L4, L5 and L6, L7 and L8.

Therefore,

$$e_{total} = p^h \times S\_W_h^n \times s'/(\tau^h \times H1) + i^h \times (\frac{5T}{28} \times H1 - S\_W_h^n \times s'/(\tau^h \times H1) + p^h \times S\_H_h^n \times s'/(\tau^h \times H2) + i^h \times (\frac{2T}{28} \times H2 - S\_H_h^n \times s'/(\tau^h \times H2) + p^h \times M\_W_h^n \times s'/(\tau^h \times H3) + i^h \times (\frac{5T}{28} \times H3 - M\_W_h^n \times s'/(\tau^h \times H3) + p^h \times M\_H_h^n \times s'/(\tau^h \times H4) + i^h \times (\frac{2T}{28} \times H4 - M\_H_h^n \times s'/(\tau^h \times H4) + p^h \times A\_W_h^n \times s'/(\tau^h \times H5) + i^h \times (\frac{5T}{28} \times H5 - A\_W_h^n \times s'/(\tau^h \times H5) + p^h \times A\_H_h^n \times s'/(\tau^h \times H6) + i^h \times (\frac{2T}{28} \times H6 - A\_H_h^n \times s'/(\tau^h \times H6) + p^h \times W\_W_h^n \times s'/(\tau^h \times H7) + i^h \times (\frac{5T}{28} \times H7 - W\_W_h^n \times s'/(\tau^h \times H7) + p^h \times W\_H_h^n \times s'/(\tau^h \times H8) + i^h \times (\frac{2T}{28} \times H8 - W\_H_h^n \times s'/(\tau^h \times H8) + p^l \times S\_W_l^n \times s'/(\tau^l \times L1) + i^l \times (\frac{5T}{28} \times L1 - S\_W_l^n \times s'/(\tau^l \times L1)) + p^l \times S\_H_l^n \times s'/(\tau^l \times L2) + i^l \times (\frac{2T}{28} \times L2 - S\_H_l^n \times s'/(\tau^l \times L2)) + p^l \times M\_W_l^n \times s'/(\tau^l \times L3) + i^l \times (\frac{5T}{28} \times L3 -$$

$$M\_W_l^n \times s'/(\tau^l \times L3))+p^l \times M\_H_l^n \times s'/(\tau^l \times L4) + i^l \times (\frac{2T}{28} \times L4 - M\_H_l^n \times s'/(\tau^l \times L4))+p^l \times A\_W_l^n \times s'/(\tau^l \times L5)+i^l \times (\frac{5T}{28} \times L5 - A\_W_l^n \times s'/(\tau^l \times L5))+p^l \times A\_H_l^n \times s'/(\tau^l \times L6) + i^l \times (\frac{2T}{28} \times L6 - A\_H_l^n \times s'/(\tau^l \times L6))+p^l \times W\_W_l^n \times s'/(\tau^l \times L7)+i^l \times (\frac{5T}{28} \times L7 - W\_W_l^n \times s'/(\tau^l \times L7))+p^l \times W\_H_l^n \times s'/(\tau^l \times L8) + i^l \times (\frac{2T}{28} \times L8 - W\_H_l^n \times s'/(\tau^l \times L8)) \quad (18)$$

Assume $B=p^h \times s'/\tau^h$ $C=p^l \times \times s'/\tau^l$

The Energy Consumption Model can be simplified as :

$$e_{total}= ({S_W}_h^n/H1 + {S_H}_h^n/H2 + {M_W}_h^n/H3 + {M_H}_h^n/H4 + {A_W}_h^n/H5 + {A_H}_h^n/H6 + {W_W}_h^n/H7 + {W_H}_h^n/H8) \times B+i^h \times (\frac{5T}{28} \times H1 - S\_W_h^n \times s'/(\tau^h \times H1)+i^h \times (\frac{2T}{28} \times H2 - S\_H_h^n \times s'/(\tau^h \times H2) + i^h \times (\frac{5T}{28} \times H3 - M\_W_h^n \times s'/(\tau^h \times H3)+ i^h \times (\frac{2T}{28} \times H4 - M\_H_h^n \times s'/(\tau^h \times H4) + i^h \times (\frac{5T}{28} \times H5 - A\_W_h^n \times s'/(\tau^h \times H5)+ i^h \times (\frac{2T}{28} \times H6 - A\_H_h^n \times s'/(\tau^h \times H6) + i^h \times (\frac{5T}{28} \times H7 - W\_W_h^n \times s'/(\tau^h \times H7)+ i^h \times (\frac{2T}{28} \times H8 - W\_H_h^n \times s'/(\tau^h \times H8) +({S_W}_l^n/L1 + {S_H}_l^n/L2 + {M_W}_l^n/L3 + {M_H}_l^n/L4 + {A_W}_l^n/L5 + {A_H}_l^n/L6 + {W_W}_l^n/L7 + {W_H}_l^n/L8) \times C +i^l \times (\frac{5T}{28} \times L1 - S\_W_l^n \times s'/(\tau^l \times L1)) + i^l \times (\frac{2T}{28} \times L2 - S\_H_l^n \times s'/(\tau^l \times L2))+i^l \times (\frac{5T}{28} \times L3 - M\_W_l^n \times s'/(\tau^l \times L3)) + i^l \times (\frac{2T}{28} \times L4 - M\_H_l^n \times s'/(\tau^l \times L4))+i^l \times (\frac{5T}{28} \times L5 - A\_W_l^n \times s'/(\tau^l \times L5)) + i^l \times (\frac{2T}{28} \times L6 - A\_H_l^n \times s'/(\tau^l \times L6))+i^l \times (\frac{5T}{28} \times L7 - W\_W_l^n \times s'/(\tau^l \times L7))+i^l \times (\frac{2T}{28} \times L8 - W\_H_l^n \times s'/(\tau^l \times L8)) \quad (19)$$

Correspondingly, the energy consumption model of the Hadoop default data placement strategy is as follows, which is without classification:

$$e'_{total} = e'^{active}_h + e'^{idle}_h$$

$$e'^{active}_h = p^h \times ({S_W}_h^n + {S_H}_h^n+{M_W}_h^n+{M_H}_h^n + {A_W}_h^n + {A_H}_h^n + {W_W}_h^n + {W_H}_h^n + S\_W_l^n + S\_H_l^n + M\_W_l^n + M\_H_l^n + A\_W_l^n + A\_H_l^n + W\_W_l^n + W\_H_l^n) \times s'/(\tau^h \times n)$$

$$=({S_W}_h^n + {S_H}_h^n+{M_W}_h^n+{M_H}_h^n + {A_W}_h^n + {A_H}_h^n + {W_W}_h^n + {W_H}_h^n + {S_W}_l^n + {S_H}_l^n + {M_W}_l^n + {M_H}_l^n + {A_W}_l^n + {A_H}_l^n + {W_W}_l^n + {W_H}_l^n) \times B/n$$

$$e'^{idle}_h = i^h \times (T \times n - ({S_W}_h^n + {S_H}_h^n+{M_W}_h^n+{M_H}_h^n + {A_W}_h^n + {A_H}_h^n + {W_W}_h^n + {W_H}_h^n + S\_W_l^n + S\_H_l^n + M\_W_l^n + M\_H_l^n + A\_W_l^n + A\_H_l^n + W\_W_l^n + W\_H_l^n) \times s'/(\tau^h \times n)$$

Therefore, $e'_{total} = ({S_W}_h^n + {S_H}_h^n+{M_W}_h^n+{M_H}_h^n + {A_W}_h^n + {A_H}_h^n + {W_W}_h^n + {W_H}_h^n + {S_W}_l^n + {S_H}_l^n + {M_W}_l^n + {M_H}_l^n + {A_W}_l^n + {A_H}_l^n + {W_W}_l^n + {W_H}_l^n) \times B/n +i^h \times (T \times n - ({S_W}_h^n + {S_H}_h^n+{M_W}_h^n+{M_H}_h^n + {A_W}_h^n + {A_H}_h^n + {W_W}_h^n + {W_H}_h^n + S\_W_l^n + S\_H_l^n + M\_W_l^n + M\_H_l^n + A\_W_l^n + A\_H_l^n + W\_W_l^n + W\_H_l^n) \times s'/(\tau^h \times n)$ (20)

And the Energy Consumption model of SEA classification [12] can be deduced by the following formula:

$e_{total_{sea}} = e_{\mathrm{hot}} + e_{cold} = e_{hot}^{active} + e_{hot}^{idle} + e_{cold}^{active} + e_{cold}^{idle} = \mathrm{p}^h \times n_h \times \mathrm{s}'/(\tau^h \times \eta_h \times n) + i^h \times (T \times \eta_h \times n - n_h \times \mathrm{s}'/(\tau^h \times \eta_h \times n)) + \mathrm{p}^l \times n_c \times \mathrm{s}'/(\tau^l \times \eta_c \times n) + i^l \times (T \times \eta_c \times n - n_c \times \mathrm{s}'/(\tau^l \times \eta_c \times n))$

(21)

Where η_h is the ratio of hot (popular) data and η_c is the ratio of cold (unpopular) data. n_h is the total access time of hot data, and n_c is the total access time of cold data.

As the mathematical models of the K-ear, Hadoop-default and SEA are stated in formula (19), (20), (21) respectively. We can deduce that the time complexity of the three algorithms depends on the time to calculate the access frequency of the data. That is to say, the time complexity of the three algorithms is O(n). On the other hand, space complexity of the three algorithms depends on the space to store the access frequency of the data. That is to say, the space complexity of the three algorithms is also O(n). As the analyzed above, the advantage of the three algorithms are usually depends on the energy consumption savings.

Based on the energy consumption model, simulation experiments with the K-ear algorithm and the SEA and Hadoop default algorithms will be conducted in the following section.

5. Performance Evaluation

To evaluate the energy efficiency of the proposed K-ear strategy, we generate the workload according to the real access trace from the wiki (wiki workload) and assume a disk with two speed modes. The different transfer rates and energy consumption rates are extracted from article [12]. All of the simulation experiments are conducted in CloudSimDisk [36], which builds on the CloudSim [37] simulator and adds storage simulation capabilities. And the detailed experimental parameters about the software and hardware configuration are listed in Table 7.

The general parameters utilized in the experiments are listed in Table 8.

Disks with two speed modes, the workload and the disk zone, are modeled in the CloudSimDisk environment. The energy efficiency of the three algorithms is evaluated by setting the following ratios: (1) The ratio of the high speed disk utilization to the cloud storage system utilization, (2) the ratio of the data with seasonal characteristics to the data without seasonal characteristics, (3) the ratio of the data with tidal characteristics to the data without tidal characteristics, and (4) the ratio of the hot data to the cold data.

5.1 The impact of the different ratios of high-speed disk utilization on cloud storage system utilization

The parameters corresponding to the values set in the experiment are listed in Table 9.

In Table 9, The ratios of the other seasons with the different characteristics are the same as the Spring season, with the s in the notation η_s is respectively replaced by m, a, and w.

In order to calculate the energy consumption of the SEA algorithm, the ratio of the hot data to cold data is set to 4: 6 in this experiment. Obtained experiment results are demonstrated in the following tables (Table 10, Table 11, Table 12), when the ratios of the high disk utilization to the whole system utilization are 1.6, 1.8, 2.0 respectively.

As shown in the Table 10, the ratio of the high disk utilization to the whole system utilization is 1.6, energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is more than 40%, while compared to the SEA, the saved energy consumption is about 11%. Moreover, we have found that the system utilization has little impact on the energy consumption for the three algorithms.

As shown in the Table 11, when the ratio of the high disk utilization to the whole system utilization is 1.8, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 42%, while compared to the SEA, the saved energy consumption is also about 11%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms in this experiment.

As shown in the Table 12, when the ratio of the high disk utilization to the whole system utilization is 2.0, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 42%, while compared to the SEA, the saved energy consumption is also about 11%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms in this experiment.

The above experimental results show that under the different ratios of high-speed disk utilization to system utilization, the K-ear strategy consumes the least energy and the Hadoop default strategy consumes the most energy. The amount of energy consumed by SEA is between the amounts consumed by the K-ear strategy and the Hadoop default algorithm. Moreover, the energy consumption of the cloud storage system only slightly increased as the ratio increased.

5.2 The impact of the different ratios of data with seasonal characteristics to data without seasonal characteristics

The ratio of the high-speed disk utilization to the system utilization is set to 2.0, and the ratio of hot data to cold data is set to 4:6 in all of the experiments (experiment 1, experiment 2, and experiment 3) in this subsection. The other parameter values are listed in the following tables (Table 13, Table 14).

In Table 13, the ratios of the other seasons with the different characteristics are the same as the Spring season, with the s in the notation η_s is respectively replaced by m, a, and w.

As the value set as in Table 12 and Table 13, the obtained experimental results are demonstrated in Table 15.

As shown in the Table 15, when the experimental parameters are set as in Table 12 and Table 13, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is also fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 36%, while compared to the SEA, the saved energy consumption is also about 1.3%, which is almost the same. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms in this experiment.

The obtained experimental results are demonstrated in Table 17.

As shown in the Table 17, when the experimental parameters are set as in Table 13 and Table 16, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is also fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 40%, while compared to the SEA, the saved energy consumption is also about 7%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms in this experiment.

Seasonal characteristics related parameters' value set in experiment 3 are listed in Table 18

The obtained experimental results are demonstrated in Table 19.

As shown in the Table 19, when the experimental parameters are set as in Table 12 and Table 17, the energy consumed by the SEAis least, and the most is the Hadoop-default. While the energy consumed by the proposed K-ear is fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 32%, while compared to the SEA, the energy consumed by K-ear is increased is about 4%, which has little difference. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms in this experiment.

As shown in the abovementioned experimental results, the K-ear and SEA strategies consume less energy than the Hadoop default strategy under different ratios of data with seasonal characteristics to data without seasonal characteristics. When the data has a higher ratio of seasonal characteristics of 6:4, K-ear slightly outperforms SEA in energy consumption. It can also be found that the lower the ratio of seasonal characteristics (48:52), SEA is more energy-efficient than K-ear but with little difference.

5.3 The impact of different ratios of data with tidal characteristics to data without tidal characteristics

The energy consumption of the three strategies with different ratios of data with tidal characteristics is tested in this subsection. The common parameters used in the different experiments are listed in Table 20.

And the tidal characteristics related parameters used in the first experiment are listed in Table 21.

The ratios of the other seasons with the different characteristics are the same as the Spring season, with the s in the notation η_s is respectively replaced by m, a, and w.

As the parameters of the first experiment set as the above tables, the obtained experimental results are shown in Table 22.

As shown in the Table 22, when the ratio of the data with tidal characteristics to the data without tidal characteristics is 4:6, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is also fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 40%, while compared to the SEA, the saved energy consumption is also about 7%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms.

Tidal characteristics related parameters used in the second experiment are listed in Table 23.

In Table 23, the ratios of the other seasons with the different characteristics are the same as the Spring season, with the s in the notation η_s is respectively replaced by m, a, and w.

As the parameters of the second experiment set as the above tables, the obtained experimental results are shown in Table 24.

As shown in the Table 24, when the ratio of the data with tidal characteristics to the data without tidal characteristics is 5:5, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is also fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 41%, while compared to the SEA, the saved energy consumption is also about 9%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms in this experiment.

Tidal characteristics related parameters used in the third experiment are listed in Table 25.

In Table 24, the ratios of the other seasons with the different characteristics are the same as the Spring season, with the s in the notation η_s is respectively replaced by m, a, and w.

As the parameters of the third experiment set as the above tables, the obtained experimental results are shown in Table 26.

As shown in the Table 26, when the ratio of the data with tidal characteristics to the data without tidal characteristics is 2:8, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is also fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 37%, while compared to the SEA, the saved energy consumption is also about 3.5%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms.

The results of the above three experiments demonstrate that our proposed K-ear strategy is more energy-efficient than the SEA algorithm when the ratio of data with tidal characteristics is high. Furthermore, K-ear has an energy efficiency advantage over the Hadoop default strategy whenever the ratio is high or low.

5.4 The impact of the different ratios of hot data to cold data

We set the different ratios of hot data to cold data while leaving the other parameters fixed. The energy consumption of the three strategies is evaluated according to the different ratios. The common parameters used in the following three experiments are listed in Table 27.

In Table 27, the ratios of the other seasons with the different characteristics are the same as the Spring season, with the s in the notation η_s is respectively replaced by m, a, and w.

Results obtained from simulation experiments when the ratio of hot data to cold data is set as 4:6 are shown in Table 28.

As shown in the Table 28, when the ratio of the hot data to the cold data is 4:6, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is also fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 42%, while compared to the SEA, the saved energy consumption is also about 11%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms.

Results obtained from simulation experiments when the ratio of hot data to cold data is set as 3:7 are shown in Table 29.

As shown in the Table 29, when the ratio of the hot data to the cold data is 3:7, the energy consumed by the K-ear is least, and the most is the Hadoop-default. While the energy consumed by the SEA is also fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 42%, while compared to the SEA, the saved energy consumption is also about 2%, which is almost the same. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms.

Results obtained from simulation experiments when the ratio of hot data to cold data is set as 2:8 are shown in Table 30.

As shown in the Table 30, when the ratio of the hot data to the cold data is 2:8, the energy consumed by the SEAis least, and the most is the Hadoop-default. While the energy consumed by the proposed K-ear is fall in between. Compared to the Hadoop-default, the average energy consumption saved by the K-ear is about 42%, while compared to the SEA, the energy consumed by K-ear is increased is about 7%. Moreover, we also have found that the system utilization has little impact on the energy consumption for the three algorithms in this experiment.

As shown in the above comparative experimental results, it can be seen that when the ratio of hot data to cold data is higher, the K-ear strategy performs better than the SEA algorithm. When the ratio is 4:6, K-ear has an obvious advantage over SEA. When the ratio is 3:7, the performance of K-ear and SEA are almost the same. When the ratio is 2:8, SEA slightly outperforms K-ear.

6. Conclusions and Future Work

An energy-aware data clustering strategy, K-ear, is proposed in this paper, in which the seasonal and tidal characteristics of data access are extracted thoroughly. The machine learning algorithm is applied to cluster data into different categories, and based on the categories, the data are stored in different storage zones. During the different time zones, some storage zones run in high performance mode to satisfy the performance requirement and the remaining storage zones run in low energy consumption mode to save energy consumption. To analyze and evaluate the energy efficiency of the proposed K-ear strategy, the famous classical SEA strategy and default data placement strategy in Hadoop are used for comparison, and mathematical models are constructed for the three strategies. Moreover, substantial simulation experiments are conducted in the CloudSimDisk simulator from different perspectives with different ratios. Compared with the mainstream data placement strategy (Hadoop default), K-ear and SEA are more energy efficient. The proposed K-ear strategy outperforms the classical SEA algorithm in most cases. Only when the ratio of hot data is very low is SEA more energy-efficient than K-ear. As a whole, compared with the other evaluated algorithms, the proposed K-ear data clustering storing strategies extracted the data access characteristics more thoroughly and classified the data into fine granularity categories, which will achieve more energy consumption savings with different disk running modes. Furthermore, as algorithm of SEA outperform the algorithms of Greedy [41], SP (Sort Partitions) [42], HP (Hybrid Partition), [42] and PVFS (Parallel Virtual File System)[43]. And the Hadoop default is more energy efficient than the algorithm of Datacenter without energy management. One part of our future work, we will explore how to combine the advantages of the different data placement strategies to achieve higher energy consumption reduction in cloud storage systems.

CRediT author statement

Xindong You: Conceptualization, Methodology, Validation, Writing original draft, Funding acquisition. **Tian Sun:** Formal analysis, Investigation, Data curation, Writing - review & editing. **Dawei Sun:** Validation, Investigation, Writing - review & editing. **Xueqiang Lv:** Conceptualization Supervision, Writing - review & editing, Funding acquisition. **Xunyun Liu**: Data curation, Investigation, Writing - review & editing. **Rajkumar Buyya:** Conceptualization, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grants (No. 61671070, 61972364.), National Language Committee of China under Grants ZDI135-53, Australian Research Council (ARC) Discovery Project and Project of Developing University Intension for Improving the Level of Scientific Research–No.2019KYNH226, Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University No.QXTCP B201908.

References

[1] G.S. Aujla, N.Kumar. MEnSuS: An efficient scheme for energy management with sustainability of cloud data centers in edge–cloud environment. Future Generation Computer Systems. 86(9): 1279--1300. 2018.

[2] R. Kaushik. Energy management costs for a data center. USA Patent, Current Assignee: International Business Machines Corp. Granted in 2018.

[3] B. Varghese, R. Buyya. Next generation cloud computing: New trends and research directions. Future Generation Computer Systems. 79(3): 849-861. 2018.

[4] M. Mills, "The Cloud Begins with Coal: Big Data, Big Networks, Big Infrastructure, and Big Power," in National Mining Association & American Coalition for clean Coal Electricity, 2013.

[5] A.R.Dehghani-Sanij, E.Tharumalinga, M.B Dusseault, R.Fraser. Study of Energy Storage Systems and Environmental Challenges of Batteries. Renewable and Sustainable Energy Reviews. 104 (4) : 192-208, 2019.

[6] D.Popa, F.Pop, C. Serbanescu, A. Castiglione. Deep learning model for home automation and energy reduction in a smart home environment platform. Neural Computing and Applications. 31(5) : 1317-1337. 2019.

[7] Envantage. New Report Reveals Warehouses Overspend on Energy by £190m. Data by World Resources Institute. (14th August, 2013). Retrieved September 13.

[8] Uzaman, Sardar Khaliq, et al. A systems overview of commercial data centers: initial energy and cost analysis. *International Journal of Information Technology and Web Engineering (IJITWE)* 14.1 (2019): 42-65.

[9] Marahatta, Avinab, et al. Classification-based and energy-efficient dynamic task scheduling scheme for virtualized cloud data center. IEEE Transactions on Cloud Computing (2019).

[10] Lin, Hsueh-Yuan, and Sheng-Yuan Yang. A smart cloud-based energy data mining agent using big data analysis technology. Smart Science 7.3 (2019): 175-183.

[11] Asgari, Sahar, et al. Hybrid surrogate model for online temperature and pressure predictions in data centers. Future Generation Computer Systems (2020).

[12] T. Xie, SEA: A Striping-Based Energy-Aware Strategy for Data Placement in RAID-Structured Storage Systems. IEEE Transactions on Computers, Vol.57, No.6, June 2008.

[13] R.T.Kaushik, and M. Bhandarkar. GreenHDFS: towards an energy-conserving, storage-efficient, hybrid Hadoop compute cluster, in Proc. of the 2010 international conference on Power aware computing and systems. HotPower '10, USENIX Association, Berkeley, CA,USA, 1-9, 2010.

[14] R. T.Kaushik., L.Cherkasova., R.Campbell, and R.Nahrstedt, Lightning: self-adaptive, energy-conserving, multi-zoned, commodity green cloud storage system, in Proc. of the 19th ACM International Symposium on High Performance Distributed Computing. 2010, HPDC '10.ACM, New York, NY, USA, 332-35.

[15] 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.9, No.4, 2151-2160 (2015).

[16] T. Zhang, B. Liao, H. Sun, F.G. Li, J.H. Ji. Energy-efficient algorithm based on data classification for cloud storage system. Journal of Computer Application, 2014, 34(8):2267-2273.

[17] X.L. Xu, G. Yang, L.J. Li, R.C. Wang. Dynamic data aggregation algorithm for data centers of green cloud computing. Systems Engineering and Electronics. Vol.34, No.9, September 2012.

[18] S.Q. Long. Research on Data Layout Strategies for Cloud Storage System. PHD thesis, South China University of Technology. 2014.

[19] B. Liao, J.Yu, H. Sun et al. Energy-efficient algorithms for distributed storage system based on data storage structure reconfiguration. Journal of Computer Research and Development, 2013, 50 (1) :3-18.

[20] R. Reddy, A. Kathpal, J. Basak, R. Katz. Data layout for power efficient archival storage systems. HotPower'15 Proceedings of the workshop on Power-Aware Computing and Systems. 2015.

[21] Z.Song. T.Wang,.T.Li, G. Yu. Energy Consumption Optimization Data Placement Algorithm for MapReduce System. Journal of Software, 2015, 26(8): 2091-2110.

[22] S.P. Ahuja, K. Muthiah. Survey of State-of-Art in Green Cloud Computing. International Journal of Green Computing (IJGC) 7(1). 2016.

[23] R. Yadav, Z, . "MeReg: Managing energy-SLA tradeoff for green mobile cloud computing." Wireless Communications and Mobile Computing 2017 (2017).

[24] R.Yadav, W.Zhang, H. Chen, T.Guo. (2017, August). Mums: Energy-aware vm selection scheme for cloud data center. In 2017 28th International Workshop on Database and Expert Systems Applications (DEXA) (pp. 132-136). IEEE.

[25] R.Yadav, et al. "Adaptive energy-aware algorithms for minimizing energy consumption and SLA violation in cloud computing." IEEE Access 6 (2018): 55923-55936.

[26] R.Yadav, et al. "An adaptive heuristic for managing energy consumption and overloaded hosts in a cloud data center." Wireless Networks 26.3 (2020): 1905-1919.

[27] Q.Wu, Q.Y. Deng, L. Ganesh. Dynamo: Facebook's Data Center- Wide Power Management System.2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA). 2016.

[28] R. Yadav, W.Z.Zhang, O. Kaiwartya, P.R.Singh, et al. Adaptive energy-aware algorithms for minimizing energy consumption and SLA violation in cloud computing, IEEE Access, 2018. Volume (6):55923-55936.

[29] L.Z.Wang, Y. Ma, J.N.Yan, V.Chang, A. Y. Zomaya. pipsCloud: High-performance cloud computing for remote sensing

big data management and processing, Future Generation Computer Systems, Volume 78, Part 1, 2018: 353-368.

[30] J. Song, H.Y.He, G. Yu, J.M. Pierson. Modulo Based Data Placement Algorithm for Energy Consumption Optimization of MapReduce System. Journal of Grid Computing .2018, 16(3):409-424.

[31] X.T.Tran, T.V.Do, C. Rotter, D. Hwang. A New Data Layout Scheme for Energy-Efficient MapReduce Processing Tasks. Journal of Grid Computing. 2018, 16(2): 285-298.

[32] Y.Ebadi, N.J. Navimipour. An energy-aware method for data replication in the cloud environments using a Tabu search and particle swarm optimization algorithm. Wiley Online Library, 28 August, 2018.

[33] R. Vales, J. Moura, R. Marinheiro. Energy-aware and adaptive fog storage mechanism with data replication ruled by spatio-temporal content popularity. Journal of Network and Computer Applications. 2019, 135(6): 84-96.

[34] C.L.Li, Y.P. Wang, Y.Chen, Y.L. Luo. Energy-efficient Fault-tolerant Replica Management Policy with Deadline and Budget Constraints in Edge-cloud Environment. Journal of Network and Computer Applications. 2019,143(1): 152-166.

[35] A.A. Khan, M.Zakarya, R.Khan, I. Rahman, M.Khan. An Energy, Performance Efficient Resource Consolidation Scheme for Heterogeneous Cloud Datacenters. Volume 150, 15 January 2020, 102497

[36] B. Louis, K. Mitra, S. Saguna, and C. Åhlund, CloudSimDisk: Energy-Aware Storage Simulation in CloudSim, Proceedings of the 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC), Limassol, Cyprus.

[37] R. Calheiros, R. Ranjan, A. Beloglazov, C. Rose, and R. Buyya, CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms, Software: Practice and Experience, Volume 41, Number 1, Pages: 23-50, ISSN: 0038-0644, Wiley Press, New York, USA, January 2011.

[38] S. Chatradhi. Hard Drive for Low Power Energy Efficiency in Disk Storage, © 2009 Hitachi Global Storage Technologies https://e3s-center.berkeley.edu/wp-content/uploads/2017/07/Sridhar%20Energy_efficient_hdd_June2009.pdf

[39] A. Hylick, R. Sohan, A. Rice and B. Jones, "An Analysis of Hard Drive Energy Consumption," 2008 IEEE International Symposium on Modeling, Analysis and Simulation of Computers and Telecommunication Systems, Baltimore, MD, 2008, pp. 1-10, doi: 10.1109/MASCOT.2008.4770567.

[40] J.Zedlewski, S. Sobti, N. Garg, Z. Fengzhou,et al.Modeling Hard-Disk Power Consumption Proceedings of FAST '03:2nd USENIX Conference onFile and Storage TechnologiesSan Francisco, CA, USAMarch 31–April 2, 2003.

[41] R.L. Graham, Bounds on Multiprocessing Timing Anomalies, SIAM J. Applied Math., vol. 7, no. 2, pp. 416-429, 1969.

[42] L.W. Lee, P. Scheuermann, and R. Vingralek, File Assignment in Parallel I/O Systems with Minimal Variance of Service Time, IEEE Trans. Computers, vol. 49, no. 2, pp. 127-140, Feb. 2000.

[43] R. Latham, N. miller, R. Ross, and P. Carns, A Next-Generation Parallel File System for Linux Clusters: An Introduction to the Second Parallel Virtual File System, Linux World Magazine, pp. 56-59, Jan. 2004.

Table 1 Summary of the reviewed related work compared with our work

Work	Name	Access frequency	Periodic access characteristics		Zones divided
			Seasonal characteristics	Tidal characteristics	
T. Xie [12]	SEA	✓	–	–	Hot Disk Zone, Cold Disk Zone
R.T. Kaushik et al [13]	GreenHDFS	✓	–	–	Hot Zone, Cold Zone
R.T. Kaushik et al [14]	Lightning	✓	–	–	Hot Zone, Cold Zone
X,D.You et al [15]	AGDC	✓	✓	–	Cold Zone, Seasonal Hot Zone, Hot Zone
Z. Tao et al. [16]	–	✓	–	–	HotZone, ColdZone and Reduplication Zone
X.L. Xu [17]	–	✓	–	–	Dynamically Aggregating
S.Q.Long [18]	SFLS	✓	–	–	Hot Files Zone, Big Files Zone
Our Work	K-ear	✓	✓	✓	Five Big Zones (according to seasonal characteristics) with Fifteen Small Zones (according to tidal characteristics furthermore)

Algorithm 1: Seasonal Characteristics Extracting Algorithm (SCEA)

Input: Data representation Set $D = \{d_1, d_2, \cdots, d_m\}$;
The number of the weeks y;

Output: Data Seasonal Characteristic $SE = \begin{bmatrix} Se_1 \\ Se_2 \\ \vdots \\ Se_m \end{bmatrix}$

Begin

for each data $d_i \in D$ do
  Parse the image data from Baidu index page into the Search Index Number of one week: $ZS = \{zs_1, zs_2, \cdots, zs_{y*52}\}$;
    Calculate the sum of searching index for each season;
      Initialize the sum_of_season vector to zero;
      for (j=1; j<=y*4;j++) // If there are y years, there are y*4 seasons.
          sum_of_season[j]=0
      End for
      j=1; from the first year
      for(k=1;k<=y*52;k++) //Travel every weeks to calculate the search index number of every seasons in different years
          sum_of_season[j]+=zs[k] //calculate the index number of every seasonswhile k is the times of 12.
          if the season ending, Thas is if (k%13==0)

```
                go to next season j++;
            End if
        End for
    Calculate the sum of searching index for each year
        Initialize the sum_of_year vector to zero.
        for (j=1; j<=y;j++)
            sum_of_year[j]=0
        End for
        j=1; from the first season
        for(k=1;k<=y*4;k++) // travel the seasons to calculate the search index of every years.
            sum_of_year[j]+=sum_of_season[k]
            if the year ending, Thas is if (k%4==0)
                go to the next season j++;
            End if
        End for
    Calculate the frequency of searching index for each season
      Initialize the frequency_of_season vector to zero
    for(k=1;k<=y*4;k++)
        se[k]=0.0;
    End for
    for(k=1;k<=y*4;k++)
      se[k]=sum_of_season[k]/sum_of_year[k/4+1]// Calculate the frequency of searching index for each season
    End for
End for
End
```

Algorithm 2: Tidal Characteristics Extracting Algorithm (TCEA)

Input: Data representation Set $D = \{d_1, d_2, \cdots, d_m\}$;
The number of the weeks z;

Output: Data Tidal Characteristics $CX = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_m \end{bmatrix}$

Begin

for each data $d_i \in D$ do
Parse the image data from Baidu index page to the Search Index Number of one day: $S = \{s_1, s_2, \cdots, s_{z*7}\}$;
Find the peak and the valley value of every week;
Initialize the first week k=1 and the index of first week's tidal value x=1;
for j=1; j<=z*7;j++
Record the index of the weak peak and valley value

```
        Initial value is: p_index=k;v_index=k;
         if (s[j]>s[p_index]) p_index=j; // Record the index of the weak peak
         End if
          if (s[j]<s[v_index]) v_index=j; // Record the index of the valley value
         End if
         if( j%7==0) //new week will begin, record the week tidal value
           if peak value is on the working days (from Monday to Thursday)
             That is (p_index-k)%7=1or 2 or 3 or 4   cx_i[x] = 1
           End if
           if peak value is on the weekends (Saturday or Sunday)
             That is (p_index-k)%7=0 or 6   cx_i[x] = 2
           End if
           if peak value is on Friday
             That is (p_index-k)%7=5   cx_i[x] = 3
           End if
           if valley value is on the working days (from Monday to Thursday)
             That is (p_index-k)%7=1or 2 or 3 or 4   cx_i[x + 1] = −1
           End if
           if valley value is on the weekends (Saturday or Sunday)
             That is (p_index-k)%7=0 or 6   cx_i[x + 1] = −2
           End if
           if valley value is on Friday
             That is (p_index-k)%7=5   cx_i[x + 1] = −3
           End if
        End if
        Reset the initial peak and valley value index of the new week k=j+7;
        Set the index of the new week's tidal value x=x+2;
      End for
End for
End
```

Table 2: Symbols in Algorithms related to the seasonal data proportion

Symbol	Meaning	Symbol	Meaning
η_s	Ratio of data with spring characteristics	η_{s-w}	Ratio of data with Spring and work characteristics
		η_{s-h}	Ratio of data with Spring and entertainment characteristics
		η_{s-o}	Ratio of data with Spring but without characteristics
η_o	Ratio of data without seasonal characteristics	η_{o-w}	Ratio of data with work but without seasonal characteristics
		η_{o-h}	Ratio of data with entertainment but without seasonal characteristics
		η_{o-o}	Ratio of data without seasonal and tidal characteristics

Table 3: The number of disks to store the data with tidal characteristics of every seasonal zone

Parameters	Meanings
$\eta_{s-\mathrm{w}} \times n_1$	The number of disks to store the data with spring and work characteristics
$\eta_{s-\mathrm{h}} \times n_1$	The number of disks to store the data with spring and entertainment characteristics
$\eta_{s-\mathrm{o}} \times n_1$	The number of disks to store the data with spring but without tidal characteristics
$\eta_{o-\mathrm{w}} \times n_5$	The number of disks to store the data with work but without seasonal characteristics
$\eta_{o-\mathrm{h}} \times n_5$	The number of disks to store the data with entertainment but without seasonal characteristics
$\eta_{o-\mathrm{o}} \times n_5$	The number of disks to store the data without work seasonal characteristics

Algorithm 3: Energy-aware data classification based on K-means (K-ear)

Input: Data related characteristics: a collection of *m* data in the set D, and the corresponding parameters: η_s, η_m, η_a, η_w, η_o and $\eta_{s-\mathrm{w}}$, $\eta_{s-\mathrm{h}}$, $\eta_{s-\mathrm{o}}$, $\eta_{m-\mathrm{w}}$, $\eta_{m-\mathrm{h}}$, $\eta_{m-\mathrm{o}}$, $\eta_{a-\mathrm{w}}$, $\eta_{a-\mathrm{h}}$, $\eta_{\mathrm{a-o}}$, $\eta_{w-\mathrm{w}}$, $\eta_{w-\mathrm{h}}$, $\eta_{w-\mathrm{o}}$, $\eta_{o-\mathrm{w}}$, $\eta_{o-\mathrm{h}}$, $\eta_{o-\mathrm{o}}$.

The exacted data characteristics: the output of the algorithm1 and algorithm 3: Data Tidal Characteristics $\mathrm{CX} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_m \end{bmatrix}$, and the Data Seasonal Characteristics $\mathrm{SE} = \begin{bmatrix} \mathrm{Se}_1 \\ \mathrm{Se}_2 \\ \vdots \\ \mathrm{Se}_m \end{bmatrix}$

Disk related characteristics: A disk array *DISK* with *n* size, every disk with 2-speed mode.

Output: Allocation of the data on the corresponding storage zones.

Begin

1: **for** each data $\mathrm{d}_i \in D$ do

2: Based on the Data Seasonal Characteristics $\mathrm{SE} = \begin{bmatrix} \mathrm{Se}_1 \\ \mathrm{Se}_2 \\ \vdots \\ \mathrm{Se}_m \end{bmatrix}$, and use the K-means clustering algorithm to classify the data into five classes firstly;

3: **if** d_i has the spring characteristics (**Class 1**)

4: place it evenly into the following storage zone $\mathrm{S} = \{\mathrm{s}_1, \mathrm{s}_2, \cdots, \mathrm{s}_{n1}\}$ **(Storage Zone 1)**

5: based on the Data Tidal Characteristics $\mathrm{CX}_1 = \begin{bmatrix} cx_{11} \\ cx_{12} \\ \vdots \\ cx_{1\eta_s \times m} \end{bmatrix}$, and use the K-means clustering algorithm to classify the data into three classes;

If d_i has the working day characteristics (**Sub_Class 1-1**)

place the data into the $1 \sim \eta_{s-w} \times n_1$ disks evenly(**Storage Zone 1-1**) // disks numbered by ordering

End if

If d_i has the holiday characteristics(**Sub_Class 1-2)**

place the data into the $\eta_{s-w} \times n_1 + 1 \sim \eta_{s-w} \times n_1 + \eta_{s-h} \times n_1$ disks evenly; **(Storage Zone 1-2)** // disks numbered by ordering

End if

If d_i has no working and holiday characteristics (**Sub_Class 1-3**)

place the data into the $\eta_{s-w} \times n_1 + \eta_{s-h} \times n_1 + 1 \sim n_1$ disks evenly **(Storage Zone 1-3)** // disks numbered by ordering

End if

End if

if d_i has the summer characteristics (**Class 2**)

place it evenly into the following storage zone $M = \{m_1, m_2, \cdots m_{n2}\}$; **(Storage Zone 2)**

based on the Data Tidal Characteristics $CX_2 = \begin{bmatrix} cx_{21} \\ cx_{22} \\ \vdots \\ cx_{2\eta_m \times m} \end{bmatrix}$, and use the K-means clustering algorithm to classify the data into three classes;

If d_i has the working day characteristics (**Sub_Class 2-1**)

place the data into the $1 \sim \eta_{m-w} \times n_2$ disks evenly; // disks numbered by ordering **(Storage Zone 2-1)**

End if

If d_i has the holiday characteristics (**Sub_Class 2-2**)

place the data into the $\eta_{m-w} \times n_2 + 1 \sim \eta_{m-w} \times n_2 + \eta_{m-h} \times n_2$ disks evenly; **(Storage Zone 2-2)** // disks numbered by ordering

End if

If d_i has no working and holiday characteristics (**Sub_Class 2-3**)

place the data into the $\eta_{m-w} \times n_2 + \eta_{m-h} \times n_2 + 1 \sim n_2$ disks evenly; // disks numbered by ordering **(Storage Zone 2-3)**

End if

End if

if d_i has the autumn characteristics (**Class 3**)

place it evenly into the following storage zone $A = \{a_1, a_2, \cdots a_{n3}\}$; **(Storage Zone 3)** based on the Data Tidal Characteristics $\mathrm{CX}_3 = \begin{bmatrix} cx_{31} \\ cx_{32} \\ \vdots \\ cx_{3\eta_a \times m} \end{bmatrix}$, and use the K-means clustering algorithm to classify the data into three classes;

If d_i has the working day characteristics (**Sub_Class 3-1**)

place the data into the $1 \sim \eta_{a-\mathrm{w}} \times n_3$ disks evenly; // disks numbered by ordering **(Storage Zone 3-1)**

End if

If d_i has the holiday characteristics (**Sub_Class 3-2**)

place the data into the $\eta_{a-\mathrm{w}} \times n_3 + 1 \sim \eta_{a-\mathrm{w}} \times n_3 + \eta_{s-\mathrm{h}} \times n_3$ disks evenly; **(Storage Zone 3-2)** // disks numbered by ordering

End if

If d_i has no working and holiday characteristics (**Sub_Class 3-3**)

place the data into the $\eta_{a-\mathrm{w}} \times n_3 + \eta_{s-\mathrm{h}} \times n_3 + 1 \sim n_3$ disks evenly // disks numbered by ordering **(Storage Zone 3-3)**

End if

End if

if d_i has winter characteristics; (**Class 4**)

place it evenly into the following storage zone $W = \{ w_1, w_2, \cdots, w_{n4} \}$ **(Storage Zone 4)**

based on the Data Tidal Characteristics $\mathrm{CX}_4 = \begin{bmatrix} cx_{41} \\ cx_{42} \\ \vdots \\ cx_{4\eta_w \times m} \end{bmatrix}$, and use the K-means clustering algorithm to classify the data into three classes;

If d_i has the working day characteristics (**Sub_Class 4-1**)

place the data into the $1 \sim \eta_{w-\mathrm{w}} \times n_4$ disks evenly // disks numbered by ordering; **(Storage Zone 4-1)**

End if

If d_i has the holiday characteristics (**Sub_Class 4-2**)

place the data into the $\eta_{w-\mathrm{w}} \times n_4 + 1 \sim \eta_{w-\mathrm{w}} \times n_4 + \eta_{w-\mathrm{h}} \times n_4$ disks evenly; **(Storage Zone 4-2)** // disks numbered by ordering

End if

If d_i has no working and holiday characteristics (**Sub_Class 4-3**)

place the data into the $\eta_{w-\mathrm{w}} \times n_4 + \eta_{w-\mathrm{h}} \times n_4 + 1 \sim n_4$ disks evenly // disks numbered by ordering **(Storage Zone 4-3)**

End if

End if

if d_i has no seasonal characteristics (**Class 5**)
place it evenly into the following storage zone $O = \{o_1, o_2, \cdots o_{n5}\}$;
(Storage Zone 5)
based on the Data Tidal Characteristics $CX_5 = \begin{bmatrix} cx_{51} \\ cx_{52} \\ \vdots \\ cx_{5\eta_o \times m} \end{bmatrix}$, and use the K-means clustering algorithm to classify the data into three classes;
If d_i has the working day characteristics (**Sub_Class 5-1**)
place the data into the $1 \sim \eta_{o-w} \times n_5$ disks evenly; // disks numbered by ordering
(Storage Zone 5-1)
End if
If d_i has the holiday characteristics (**Sub_Class 5-2**)
place the data into the $\eta_{o-w} \times n_5 + 1 \sim \eta_{o-w} \times n_5 + \eta_{o-h} \times n_5$ disks evenly; **(Storage Zone 5-2)**
// disks numbered by ordering
End if
If d_i has no working and holiday characteristics (**Sub_Class 5-3**)
place the data into the $\eta_{o-w} \times n_5 + \eta_{o-h} \times n_5 + 1 \sim n_5$ disks evenly; // disks numbered by ordering
(Storage Zone 5-3)
End if
End if
End for
End

Table 4 The number of disks running in different modes during different time zones

Time Zone	The number of disks running in high mode	The number of disks running in low mode
Workdays in Spring	$H1 = \eta_{s-w} \times n_1 + \eta_{s-o} \times n_1 + \eta_{s-w} \times n_5 + \eta_{s-o} \times n_5$	L1=$\eta_{s-h} \times n_1$+n_2+n_3+n_4+$\eta_{s-h} \times n_5$
Weekends in Spring	$H2 = \eta_{s-h} \times n_1 + \eta_{s-o} \times n_1 + \eta_{s-h} \times n_5 + \eta_{s-o} \times n_5$	L2=$\eta_{s-w} \times n_1$+n_2+n_3+n_4+$\eta_{s-w} \times n_5$
Workdays in Summer	$H3 = \eta_{s-w} \times n_2 + \eta_{s-o} \times n_2 + \eta_{s-w} \times n_5 + \eta_{s-o} \times n_5$	L3=$\eta_{s-h} \times n_2$+n_1+n_3+n_4+$\eta_{s-h} \times n_5$
Weekends in Summer	$H4 = \eta_{s-h} \times n_2 + \eta_{s-o} \times n_2 + \eta_{s-h} \times n_5 + \eta_{s-o} \times n_5$	L3=$\eta_{s-w} \times n_2$+n_1+n_3+n_4+$\eta_{s-w} \times n_5$
Workdays in Autumn	$H5 = \eta_{s-w} \times n_3 + \eta_{s-o} \times n_3 + \eta_{s-w} \times n_5 + \eta_{s-o} \times n_5$	L5=$\eta_{s-h} \times n_3$+n_1+n_2+n_4+$\eta_{s-h} \times n_5$

Weekends in Autumn	$H6 = \eta_{s-h} \times n_3 + \eta_{s-o} \times n_3 + \eta_{s-h} \times n_5 + \eta_{s-o} \times n_5$	$L6 = \eta_{s-w} \times n_3 + n_1 + n_2 + n_4 + \eta_{s-w} \times n_5$
Workdays in Winter	$H7 = \eta_{s-w} \times n_4 + \eta_{s-o} \times n_4 + \eta_{s-w} \times n_5 + \eta_{s-o} \times n_5$	$L7 = \eta_{s-h} \times n_4 + n_1 + n_2 + n_3 + \eta_{s-h} \times n_5$
Weekends in Winter	$H8 = \eta_{s-h} \times n_4 + \eta_{s-o} \times n_4 + \eta_{s-h} \times n_5 + \eta_{s-o} \times n_5$	$L8 = \eta_{s-w} \times n_4 + n_1 + n_2 + n_3 + \eta_{s-w} \times n_5$

Table 5: Meaning of the parameters of the whole system in mathematic models

Paramter	Meaning
τ^h	Transfer rate of disks running in high mode (MB/second)
p^h	Energy consuming rate of disks running in high mode during active status (J/second)
i^h	Energy consuming rate of disks running in high mode during idle status (J/second)
τ^l	Transfer rate of disks running in low mode (MB/second)
p^l	Energy consuming rate of disks running in low mode during active status (J/second)
i^l	Energy consuming rate of disks running in low mode during idle status (J/second)
s'	The average size of the dataset (MB)
e_{total}	The total enery consumption (J)
e_h	Energy consumed by all of the disks running in high mode (J)
e_l	Energy consumed by all of the disks running in low mode (J)
e_h^{active}	Energy consumed by all of the disks running in high mode during active status (J)
e_h^{idle}	Energy consumed by all of the disks running in high mode during idle status (J)
e_l^{active}	Energy consumed by all of the disks running in low mode during active status (J)
e_l^{idle}	Energy consumed by all of the disks running in low mode during idle status (J)
t_h^{active}	Total time of disks running in high mode with active status (Second)
t_h^{idle}	Total time of disks running in high mode with idle status (Second)
t_l^{active}	Total time of disks running in low mode with active status (Second)
t_l^{idle}	Total time of disks running in low mode with idle status (Second)
T	Total service time of a disk (Second /disk)

Table 6: Meaning of the parameters about the spring season in mathematic models

$spring\_work\_t_h^{active}$	Total time of disks running in high mode with active status during workdays in Spring (Second)
$spring\_holiday\_t_h^{active}$	Total time of disks running in high mode with active status during weekends in Spring (Second)
$spring\_work\_t_h^{idle}$	Total time of disks running in high mode with idle status during workdays in Spring (Second)
$spring\_holiday\_t_h^{idle}$	Total time of disks running in high mode with idle status during weekends in Spring (Second)
$spring\_work\_t_l^{active}$	Total time of disks running in low mode with active status during workdays in Spring

	(Second)
$spring\_holiday\_t_l^{active}$	Total time of disks running in low mode with active status during weekends in Spring (Second)
$spring\_work\_t_l^{idle}$	Total time of disks running in low mode with idle status during workdays in Spring (Second)
$spring\_holiday\_t_l^{idle}$	Total time of disks running in low mode with idle status during weekends in Spring (Second)
$spring\_work\_\mathrm{e}_{total}$	Energy consumption during workdays in Spring (J)
$spring\_holiday\_\mathrm{e}_{total}$	Energy consumption during weekends in Spring (J)
$spring\_work\_e_h^{active}$	Energy Consumption of disks running in high mode with active status during workdays in Spring (J)
$spring\_holiday\_e_h^{active}$	Energy Consumption of disks running in high mode with active status during weekends in Spring (J)
$spring\_work\_e_l^{active}$	Energy Consumption of disks running in low mode with active status during workdays in Spring (J)
$spring\_holiday\_e_l^{active}$	Energy Consumption of disks running in low mode with active status during weekends in Spring (J)
$spring\_work\_e_h^{idle}$	Energy Consumption of disks running in high mode with idle status during workdays in Spring (J)
$spring\_holiday\_e_h^{idle}$	Energy Consumption of disks running in high mode with idle status during weekends in Spring (J)
$spring\_work\_e_l^{idle}$	Energy Consumption of disks running in low mode with idle status during workdays in Summer (J)
$spring\_holiday\_e_l^{idle}$	Energy Consumption of disks running in low mode with idle status during weekends in Summer (J)
$spring\_work\_T$	Total service time of a disk during workdays in Spring (Second /disk) = 5T/28
$spring\_holiday\_T$	Total service time of a disk during weekends in Spring (Second /disk) = 2T/28
$S\_W_h^n$	Total access times in high mode during workdays in Spring
$S\_W_l^n$	Total access times in low mode during workdays in Spring
$S\_H_h^n$	Total access times in high mode during weekends in Spring
$S\_H_l^n$	Total access times in low mode during weekends in Spring

Table7 Software and Hardware Configuration in the Experiments

Equipment/Software	type/version
CPU	Intel(R) Core(TM) i5-4590 CPU @ 3.30GHz 3.3 GHz
Memory Size	4.0GB
Hard Disk	1TB(TOSHIBA DT01ACA100 ATA Device)

Network Card	Realtek PCIe GBE Family Controller
Operating System	Windows 10
Energy-aware Disk Simulator	CloudSimDisk1.0
Cloud Environment Simulator	CloudSim 4.0
Programming Platform	Eclipse-Java-Luna-SR2

Table 8: General parameters value in the simulation experiments

Paramter	Value	Paramter	Value
p^h	30.26 J/Second	i^l	2.17 J/Second
i^h	5.26 J/Second	τ^l	9.3Mb/Second
τ^h	31Mb/Second	n	1000
p^l	21.33 J/Second	T	31536000 Second

Table 9 Parameters and values set in this experiment

Parameter	value	Parameter	value
Ratio of the data with Spring characteristics η_s	0.2	Ratio of data with Spring and workday characteristics η_{s-w}	0.3
		Ratio of data with Spring and weekend characteristics η_{s-h}	0.3
		Ratio of data with Spring characteristics but without tidal characteristics η_{s-o}	0.4
Ratio of the data without seasonal characteristics η_o	0.2	Ratio of data with workday characteristics but without seasonal characteristics η_{o-w}	0.3
		Ratio of data with weekend characteristics but without seasonal characteristics η_{o-h}	0.3
		Ratio of data without seasonal and tidal characteristics η_{o-o}	0.4

Table 10: Energy consumption of the three data placement strategies when the ratio of the high disk utilization to the whole system utilization is 1.6

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	95998.62547	165958.2	107658.606
0.11	96026.6813	165966.084	107683.3049
0.12	96054.73713	165973.968	107708.0039
0.13	96082.79295	165981.852	107732.7029
0.14	96110.84878	165989.736	107757.4019
0.15	96138.90461	165997.62	107782.1009
0.16	96166.96043	166005.504	107806.7999

0.17	96195.01626	166013.388	107831.4989
0.18	96223.07209	166021.272	107856.1979
0.19	96251.12792	166029.156	107880.8969
0.2	96279.18374	166037.04	107905.5959
0.21	96307.23957	166044.924	107930.2949
0.22	96335.2954	166052.808	107954.9939
0.23	96363.35123	166060.692	107979.6929
0.24	96391.40705	166068.576	108004.3919
0.25	96419.46288	166076.46	108029.0909
0.26	96447.51871	166084.344	108053.7899
0.27	96475.57453	166092.228	108078.4889
0.28	96503.63036	166100.112	108103.1879
0.29	96531.68619	166107.996	108127.8869
0.3	96559.74202	166115.88	108152.5859
0.31	96587.79784	166123.764	108177.2849
0.32	96615.85367	166131.648	108201.9838
0.33	96643.9095	166139.532	108226.6828
0.34	96671.96532	166147.416	108251.3818
0.35	96700.02115	166155.3	108276.0808
0.36	96728.07698	166163.184	108300.7798
0.37	96756.13281	166171.068	108325.4788
0.38	96784.18863	166178.952	108350.1778
0.39	96812.24446	166186.836	108374.8768

Table 11: Energy consumption of the three data placement strategies when the ratio of the high disk utilization to the whole system utilization is 1.8

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	95998.72826	165958.2	107647.5193
0.11	96026.79436	165966.084	107671.1096
0.12	96054.86047	165973.968	107694.7
0.13	96082.92657	165981.852	107718.2903
0.14	96110.99268	165989.736	107741.8806
0.15	96139.05878	165997.62	107765.4709
0.16	96167.12489	166005.504	107789.0613
0.17	96195.19099	166013.388	107812.6516
0.18	96223.2571	166021.272	107836.2419
0.19	96251.32321	166029.156	107859.8323

0.2	96279.38931	166037.04	107883.4226
0.21	96307.45542	166044.924	107907.0129
0.22	96335.52152	166052.808	107930.6033
0.23	96363.58763	166060.692	107954.1936
0.24	96391.65373	166068.576	107977.7839
0.25	96419.71984	166076.46	108001.3742
0.26	96447.78595	166084.344	108024.9646
0.27	96475.85205	166092.228	108048.5549
0.28	96503.91816	166100.112	108072.1452
0.29	96531.98426	166107.996	108095.7356
0.3	96560.05037	166115.88	108119.3259
0.31	96588.11647	166123.764	108142.9162
0.32	96616.18258	166131.648	108166.5065
0.33	96644.24868	166139.532	108190.0969
0.34	96672.31479	166147.416	108213.6872
0.35	96700.3809	166155.3	108237.2775
0.36	96728.447	166163.184	108260.8679
0.37	96756.51311	166171.068	108284.4582
0.38	96784.57921	166178.952	108308.0485
0.39	96812.64532	166186.836	108331.6389

Table 12: Energy consumption of the three data placement strategies when the ratio of the high disk utilization to the whole system utilization is 2.0

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	95998.83104	165958.2	107636.4326
0.11	96026.90742	165966.084	107658.9143
0.12	96054.98381	165973.968	107681.396
0.13	96083.06019	165981.852	107703.8776
0.14	96111.13658	165989.736	107726.3593
0.15	96139.21296	165997.62	107748.841
0.16	96167.28934	166005.504	107771.3226
0.17	96195.36573	166013.388	107793.8043
0.18	96223.44211	166021.272	107816.286
0.19	96251.5185	166029.156	107838.7676

0.2	96279.59488	166037.04	107861.2493
0.21	96307.67126	166044.924	107883.7309
0.22	96335.74765	166052.808	107906.2126
0.23	96363.82403	166060.692	107928.6943
0.24	96391.90042	166068.576	107951.1759
0.25	96419.9768	166076.46	107973.6576
0.26	96448.05318	166084.344	107996.1393
0.27	96476.12957	166092.228	108018.6209
0.28	96504.20595	166100.112	108041.1026
0.29	96532.28234	166107.996	108063.5843
0.3	96560.35872	166115.88	108086.0659
0.31	96588.4351	166123.764	108108.5476
0.32	96616.51149	166131.648	108131.0292
0.33	96644.58787	166139.532	108153.5109
0.34	96672.66426	166147.416	108175.9926
0.35	96700.74064	166155.3	108198.4742
0.36	96728.81702	166163.184	108220.9559
0.37	96756.89341	166171.068	108243.4376
0.38	96784.96979	166178.952	108265.9192
0.39	96813.04618	166186.836	108288.4009

Table 13: General parameters' values for the different ratios of the data with seasonal characteristics and the data without seasonal characteristics

Parameter	Value
Ratio of the data with Spring and workday characteristics $\eta_{s-\mathrm{w}}$	0.3
Ratio of the data with Spring and weekend characteristics $\eta_{s-\mathrm{h}}$	0.3
Ratio of the data with Spring but without tidal characteristics $\eta_{s-\mathrm{o}}$	0.4
Ratio of the data with workday characteristics but without seasonal characteristics $\eta_{o-\mathrm{w}}$	0.3

Ratio of the data with weekend characteristics but without seasonal characteristics η_{o-h}	0.3
Ratio of the data without seasonal and tidal characteristics η_{o-o}	0.4

Table 14: Seasonal characteristics related parameters' value set in experiment 1

parameter	value
Ratio of the data with Spring characteristics η_s	0.15
Ratio of the data with Summer characteristics η_m	0.15
Ratio of the data with Autumn characteristics η_a	0.15
Ratio of the data with Winter characteristics η_w	0.15
Ratio of the data without seasonal characteristics η_o	0.4

Table 15: Energy consumption of the three data placement strategies in experiment 1

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	106182.9264	165958.2	107636.4326
0.11	106206.2268	165966.084	107658.9143
0.12	106229.5272	165973.968	107681.396
0.13	106252.8276	165981.852	107703.8776
0.14	106276.128	165989.736	107726.3593
0.15	106299.4285	165997.62	107748.841
0.16	106322.7289	166005.504	107771.3226
0.17	106346.0293	166013.388	107793.8043
0.18	106369.3297	166021.272	107816.286
0.19	106392.6301	166029.156	107838.7676
0.2	106415.9305	166037.04	107861.2493
0.21	106439.2309	166044.924	107883.7309
0.22	106462.5313	166052.808	107906.2126
0.23	106485.8317	166060.692	107928.6943
0.24	106509.1321	166068.576	107951.1759
0.25	106532.4325	166076.46	107973.6576
0.26	106555.7329	166084.344	107996.1393
0.27	106579.0333	166092.228	108018.6209
0.28	106602.3337	166100.112	108041.1026
0.29	106625.6341	166107.996	108063.5843
0.3	106648.9345	166115.88	108086.0659
0.31	106672.2349	166123.764	108108.5476

0.32	106695.5353	166131.648	108131.0292
0.33	106718.8357	166139.532	108153.5109
0.34	106742.1361	166147.416	108175.9926
0.35	106765.4365	166155.3	108198.4742
0.36	106788.7369	166163.184	108220.9559
0.37	106812.0373	166171.068	108243.4376
0.38	106835.3377	166178.952	108265.9192
0.39	106858.6381	166186.836	108288.4009

Table 16: Seasonal characteristics related parameters' value set in experiment 2

parameter	value
Ratio of the data with Spring characteristics η_s	0.18
Ratio of the data with Summer characteristics η_m	0.18
Ratio of the data with Autumn characteristics η_a	0.18
Ratio of the data with Winter characteristics η_w	0.18
Ratio of the data without seasonal characteristics η_o	0.28

Table 17: Energy consumption of the three data placement strategies in experiment 2

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	100074.2443	165958.2	107636.4326
0.11	100100.5878	165966.084	107658.9143
0.12	100126.9314	165973.968	107681.396
0.13	100153.2749	165981.852	107703.8776
0.14	100179.6184	165989.736	107726.3593
0.15	100205.9619	165997.62	107748.841
0.16	100232.3054	166005.504	107771.3226
0.17	100258.6489	166013.388	107793.8043
0.18	100284.9924	166021.272	107816.286
0.19	100311.3359	166029.156	107838.7676
0.2	100337.6794	166037.04	107861.2493
0.21	100364.0229	166044.924	107883.7309
0.22	100390.3664	166052.808	107906.2126
0.23	100416.7099	166060.692	107928.6943
0.24	100443.0534	166068.576	107951.1759
0.25	100469.3969	166076.46	107973.6576
0.26	100495.7404	166084.344	107996.1393

0.27	100522.0839	166092.228	108018.6209
0.28	100548.4275	166100.112	108041.1026
0.29	100574.771	166107.996	108063.5843
0.3	100601.1145	166115.88	108086.0659
0.31	100627.458	166123.764	108108.5476
0.32	100653.8015	166131.648	108131.0292
0.33	100680.145	166139.532	108153.5109
0.34	100706.4885	166147.416	108175.9926
0.35	100732.832	166155.3	108198.4742
0.36	100759.1755	166163.184	108220.9559
0.37	100785.519	166171.068	108243.4376
0.38	100811.8625	166178.952	108265.9192
0.39	100838.206	166186.836	108288.4009

Table 18: Seasonal characteristics related parameters' value set in experiment 3

parameter	value
Ratio of the data with Spring characteristics η_s	0.12
Ratio of the data with Summer characteristics η_m	0.12
Ratio of the data with Autumn characteristics η_a	0.12
Ratio of the data with Winter characteristics η_w	0.12
Ratio of the data without seasonal characteristics η_o	0.52

Table 19: Energy consumption of the three data placement strategies in experiment 3.

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	112284.6623	165958.2	107636.4326
0.11	112304.225	165966.084	107658.9143
0.12	112323.7877	165973.968	107681.396
0.13	112343.3504	165981.852	107703.8776
0.14	112362.913	165989.736	107726.3593
0.15	112382.4757	165997.62	107748.841
0.16	112402.0384	166005.504	107771.3226
0.17	112421.6011	166013.388	107793.8043
0.18	112441.1638	166021.272	107816.286
0.19	112460.7264	166029.156	107838.7676
0.2	112480.2891	166037.04	107861.2493

0.21	112499.8518	166044.924	107883.7309
0.22	112519.4145	166052.808	107906.2126
0.23	112538.9772	166060.692	107928.6943
0.24	112558.5398	166068.576	107951.1759
0.25	112578.1025	166076.46	107973.6576
0.26	112597.6652	166084.344	107996.1393
0.27	112617.2279	166092.228	108018.6209
0.28	112636.7906	166100.112	108041.1026
0.29	112656.3532	166107.996	108063.5843
0.3	112675.9159	166115.88	108086.0659
0.31	112695.4786	166123.764	108108.5476
0.32	112715.0413	166131.648	108131.0292
0.33	112734.6039	166139.532	108153.5109
0.34	112754.1666	166147.416	108175.9926
0.35	112773.7293	166155.3	108198.4742
0.36	112793.292	166163.184	108220.9559
0.37	112812.8547	166171.068	108243.4376
0.38	112832.4173	166178.952	108265.9192
0.39	112851.98	166186.836	108288.4009

Table 20: Common parameters used in the following three experiments

parameter	value
Ratio of the data with Spring characteristics η_s	0.2
Ratio of the data with Summer characteristics η_{m}	0.2
Ratio of the data with Autumn characteristics η_{a}	0.2
Ratio of the data with Winter characteristics η_{w}	0.2
Ratio of the data without seasonal characteristics η_{o}	0.2
Ratio of the high speed disk utilization to the system utilization	2.0
Ratio of the hot data to the cold data	4:6

Table 21: Tidal characteristics related parameters used in the first experiment

Parameter	value
Ratio of the data with Spring and workday characteristics η_{s-w}	0.2
Ratio of the data with Spring and weekend characteristics η_{s-h}	0.2

Ratio of the data with Spring but without tidal characteristics η_{s-o}	0.6
Ratio of the data with workday but without seasonal characteristics η_{o-w}	0.2
Ratio of the data with weekend but without seasonal characteristics η_{o-h}	0.2
Ratio of the data without seasonal and tidal characteristics η_{o-o}	0.6

Table 22: Energy consumption of the three data placement strategies when the ratio of the data with tidal characteristics to the data without tidal characteristics is 4:6

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	99880.22558	165958.2	107636.4326
0.11	99906.65646	165966.084	107658.9143
0.12	99933.08734	165973.968	107681.396
0.13	99959.51822	165981.852	107703.8776
0.14	99985.94909	165989.736	107726.3593
0.15	100012.38	165997.62	107748.841
0.16	100038.8108	166005.504	107771.3226
0.17	100065.2417	166013.388	107793.8043
0.18	100091.6726	166021.272	107816.286
0.19	100118.1035	166029.156	107838.7676
0.2	100144.5344	166037.04	107861.2493
0.21	100170.9652	166044.924	107883.7309
0.22	100197.3961	166052.808	107906.2126
0.23	100223.827	166060.692	107928.6943
0.24	100250.2579	166068.576	107951.1759
0.25	100276.6888	166076.46	107973.6576
0.26	100303.1196	166084.344	107996.1393
0.27	100329.5505	166092.228	108018.6209
0.28	100355.9814	166100.112	108041.1026
0.29	100382.4123	166107.996	108063.5843
0.3	100408.8431	166115.88	108086.0659
0.31	100435.274	166123.764	108108.5476
0.32	100461.7049	166131.648	108131.0292
0.33	100488.1358	166139.532	108153.5109
0.34	100514.5667	166147.416	108175.9926
0.35	100540.9975	166155.3	108198.4742
0.36	100567.4284	166163.184	108220.9559
0.37	100593.8593	166171.068	108243.4376

0.38	100620.2902	166178.952	108265.9192
0.39	100646.721	166186.836	108288.4009

Table 23: Tidal characteristics related parameters used in the second experiment

Parameter	Value
Ratio of the data with Spring and workday characteristics $\eta_{s-\mathrm{w}}$	0.25
Ratio of the data with Spring and weekend characteristics $\eta_{s-\mathrm{h}}$	0.25
Ratio of the data with Spring but without tidal characteristics $\eta_{s-\mathrm{o}}$	0.5
Ratio of the data with workday but without seasonal characteristics $\eta_{o-\mathrm{w}}$	0.25
Ratio of the data with weekend but without seasonal characteristics $\eta_{o-\mathrm{h}}$	0.25
Ratio of the data without seasonal and tidal characteristics $\eta_{o-\mathrm{o}}$	0.5

Table 24: Energy consumption of the three data placement strategies when the ratio of the data with tidal characteristics to the data without tidal characteristics is 5:5

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	97939.76338	165958.2	107636.4326
0.11	97967.04052	165966.084	107658.9143
0.12	97994.31766	165973.968	107681.396
0.13	98021.5948	165981.852	107703.8776
0.14	98048.87194	165989.736	107726.3593
0.15	98076.14907	165997.62	107748.841
0.16	98103.42621	166005.504	107771.3226
0.17	98130.70335	166013.388	107793.8043
0.18	98157.98049	166021.272	107816.286
0.19	98185.25763	166029.156	107838.7676
0.2	98212.53477	166037.04	107861.2493
0.21	98239.8119	166044.924	107883.7309
0.22	98267.08904	166052.808	107906.2126
0.23	98294.36618	166060.692	107928.6943
0.24	98321.64332	166068.576	107951.1759
0.25	98348.92046	166076.46	107973.6576
0.26	98376.19759	166084.344	107996.1393
0.27	98403.47473	166092.228	108018.6209
0.28	98430.75187	166100.112	108041.1026
0.29	98458.02901	166107.996	108063.5843

0.3	98485.30615	166115.88	108086.0659
0.31	98512.58329	166123.764	108108.5476
0.32	98539.86042	166131.648	108131.0292
0.33	98567.13756	166139.532	108153.5109
0.34	98594.4147	166147.416	108175.9926
0.35	98621.69184	166155.3	108198.4742
0.36	98648.96898	166163.184	108220.9559
0.37	98676.24612	166171.068	108243.4376
0.38	98703.52325	166178.952	108265.9192
0.39	98730.80039	166186.836	108288.4009

Table 25: Tidal characteristics related parameters used in the third experiment

Parameter	Value
Ratio of the data with Spring and workday characteristics η_{s-w}	0.1
Ratio of the data with Spring and weekend characteristics η_{s-h}	0.1
Ratio of the data with Spring but without tidal characteristics η_{s-o}	0.8
Ratio of the data with workday but without seasonal characteristics η_{o-w}	0.1
Ratio of the data with weekend but without seasonal characteristics η_{o-h}	0.1
Ratio of the data with Spring and workday characteristics η_{s-w}	0.8

Table 26: Energy consumption of the three data placement strategies when the ratio of the data with tidal characteristics to the data without tidal characteristics is 2:8

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	103759.5632	165958.2	107636.4326
0.11	103784.1429	165966.084	107658.9143
0.12	103808.7226	165973.968	107681.396
0.13	103833.3023	165981.852	107703.8776
0.14	103857.882	165989.736	107726.3593
0.15	103882.4617	165997.62	107748.841
0.16	103907.0413	166005.504	107771.3226
0.17	103931.621	166013.388	107793.8043
0.18	103956.2007	166021.272	107816.286
0.19	103980.7804	166029.156	107838.7676
0.2	104005.3601	166037.04	107861.2493
0.21	104029.9398	166044.924	107883.7309
0.22	104054.5194	166052.808	107906.2126
0.23	104079.0991	166060.692	107928.6943

0.24	104103.6788	166068.576	107951.1759
0.25	104128.2585	166076.46	107973.6576
0.26	104152.8382	166084.344	107996.1393
0.27	104177.4179	166092.228	108018.6209
0.28	104201.9976	166100.112	108041.1026
0.29	104226.5772	166107.996	108063.5843
0.3	104251.1569	166115.88	108086.0659
0.31	104275.7366	166123.764	108108.5476
0.32	104300.3163	166131.648	108131.0292
0.33	104324.896	166139.532	108153.5109
0.34	104349.4757	166147.416	108175.9926
0.35	104374.0553	166155.3	108198.4742
0.36	104398.635	166163.184	108220.9559
0.37	104423.2147	166171.068	108243.4376
0.38	104447.7944	166178.952	108265.9192
0.39	104472.3741	166186.836	108288.4009

Table 27 Common parameters used in the following three experiments

Parameter	Value	Parameters	Value
Ratio of the data with Spring characteristics η_s	0.2	Ratio of data with Spring and workday characteristics $\eta_{s-\mathrm{w}}$	0.3
		Ratio of data with Spring and weekend characteristics $\eta_{s-\mathrm{h}}$	0.3
		Ratio of data with Spring characteristics but without tidal characteristics $\eta_{s-\mathrm{o}}$	0.4
Ratio of the data without seasonal characteristics η_o	0.2	Ratio of data without seasonal characteristics but with workday characteristics $\eta_{o-\mathrm{w}}$	0.3
		Ratio of data without seasonal characteristics but with weekend characteristics $\eta_{o-\mathrm{h}}$	0.3
		Ratio of data without seasonal characteristics and tidal characteristics $\eta_{o-\mathrm{o}}$	0.4

Parameter	Value
Ratio of the high speed disk utilization to the system utilization	2.0

Table 28 Energy consumption of the three data placement strategies when the ratio of the hot data to the cold data is 4:6

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]

0.1	95998.83104	165958.2	107636.4326
0.11	96026.90742	165966.084	107658.9143
0.12	96054.98381	165973.968	107681.396
0.13	96083.06019	165981.852	107703.8776
0.14	96111.13658	165989.736	107726.3593
0.15	96139.21296	165997.62	107748.841
0.16	96167.28934	166005.504	107771.3226
0.17	96195.36573	166013.388	107793.8043
0.18	96223.44211	166021.272	107816.286
0.19	96251.5185	166029.156	107838.7676
0.2	96279.59488	166037.04	107861.2493
0.21	96307.67126	166044.924	107883.7309
0.22	96335.74765	166052.808	107906.2126
0.23	96363.82403	166060.692	107928.6943
0.24	96391.90042	166068.576	107951.1759
0.25	96419.9768	166076.46	107973.6576
0.26	96448.05318	166084.344	107996.1393
0.27	96476.12957	166092.228	108018.6209
0.28	96504.20595	166100.112	108041.1026
0.29	96532.28234	166107.996	108063.5843
0.3	96560.35872	166115.88	108086.0659
0.31	96588.4351	166123.764	108108.5476
0.32	96616.51149	166131.648	108131.0292
0.33	96644.58787	166139.532	108153.5109
0.34	96672.66426	166147.416	108175.9926
0.35	96700.74064	166155.3	108198.4742
0.36	96728.81702	166163.184	108220.9559
0.37	96756.89341	166171.068	108243.4376
0.38	96784.96979	166178.952	108265.9192
0.39	96813.04618	166186.836	108288.4009

Table 29 Energy consumption of the three data placement strategies when the ratio of the hot data to the cold data is 3:7

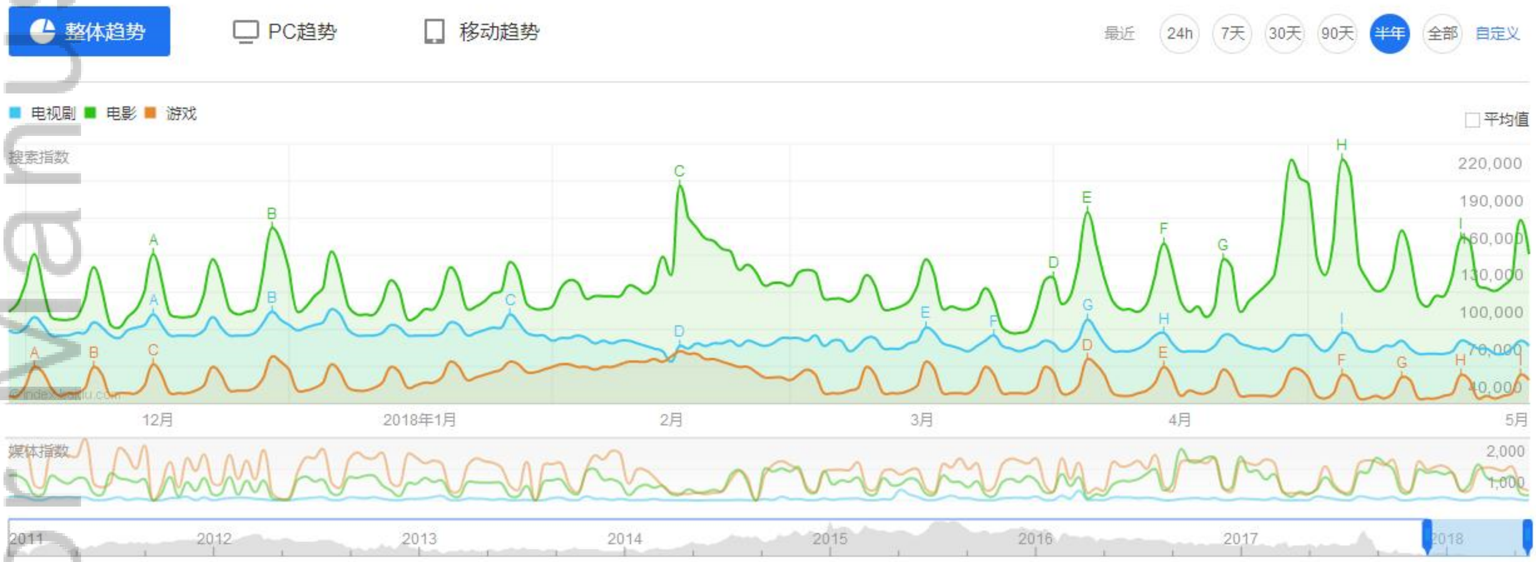
System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	95998.83104	165958.2	97939.76338
0.11	96026.90742	165966.084	97967.04052
0.12	96054.98381	165973.968	97994.31766
0.13	96083.06019	165981.852	98021.5948

0.14	96111.13658	165989.736	98048.87194
0.15	96139.21296	165997.62	98076.14907
0.16	96167.28934	166005.504	98103.42621
0.17	96195.36573	166013.388	98130.70335
0.18	96223.44211	166021.272	98157.98049
0.19	96251.5185	166029.156	98185.25763
0.2	96279.59488	166037.04	98212.53477
0.21	96307.67126	166044.924	98239.8119
0.22	96335.74765	166052.808	98267.08904
0.23	96363.82403	166060.692	98294.36618
0.24	96391.90042	166068.576	98321.64332
0.25	96419.9768	166076.46	98348.92046
0.26	96448.05318	166084.344	98376.1976
0.27	96476.12957	166092.228	98403.47473
0.28	96504.20595	166100.112	98430.75187
0.29	96532.28234	166107.996	98458.02901
0.3	96560.35872	166115.88	98485.30615
0.31	96588.4351	166123.764	98512.58329
0.32	96616.51149	166131.648	98539.86043
0.33	96644.58787	166139.532	98567.13756
0.34	96672.66426	166147.416	98594.4147
0.35	96700.74064	166155.3	98621.69184
0.36	96728.81702	166163.184	98648.96898
0.37	96756.89341	166171.068	98676.24612
0.38	96784.96979	166178.952	98703.52326
0.39	96813.04618	166186.836	98730.80039

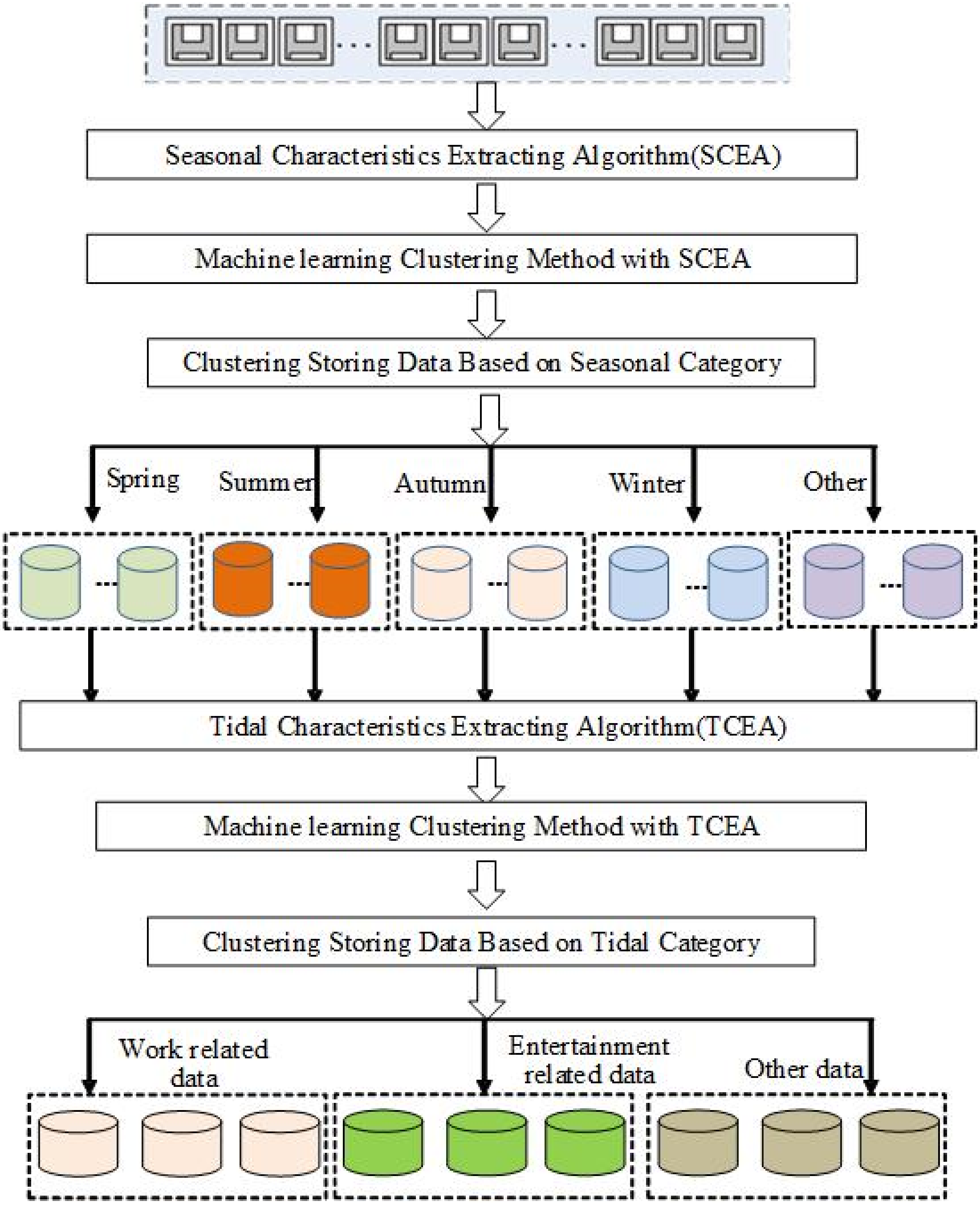
Table 30 Energy consumption of the three data placement strategies when the ratio of the hot data to the cold data is 2:8

System utilization	Energy consumption of K-ear (KJ)	Energy consumption of Hadoop-default (KJ)[13]	Energy consumption of SEA(KJ) [12]
0.1	95998.83104	165958.2	88231.10544
0.11	96026.90742	165966.084	88261.97918
0.12	96054.98381	165973.968	88292.85293
0.13	96083.06019	165981.852	88323.72667
0.14	96111.13658	165989.736	88354.60042
0.15	96139.21296	165997.62	88385.47416
0.16	96167.28934	166005.504	88416.3479
0.17	96195.36573	166013.388	88447.22165

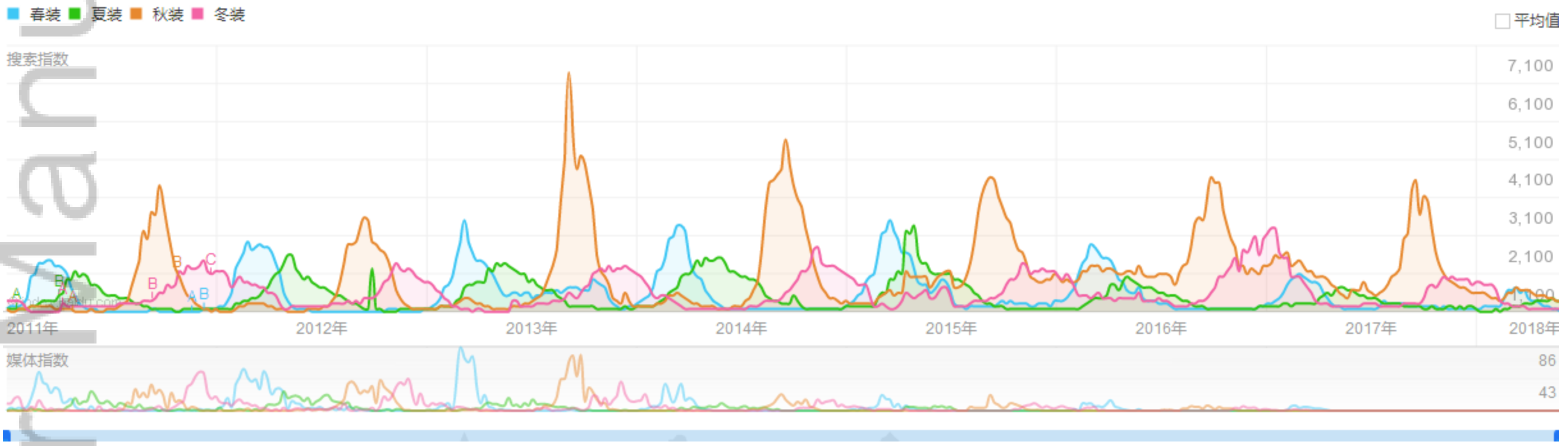
0.18	96223.44211	166021.272	88478.09539
0.19	96251.5185	166029.156	88508.96914
0.2	96279.59488	166037.04	88539.84288
0.21	96307.67126	166044.924	88570.71662
0.22	96335.74765	166052.808	88601.59037
0.23	96363.82403	166060.692	88632.46411
0.24	96391.90042	166068.576	88663.33786
0.25	96419.9768	166076.46	88694.2116
0.26	96448.05318	166084.344	88725.08534
0.27	96476.12957	166092.228	88755.95909
0.28	96504.20595	166100.112	88786.83283
0.29	96532.28234	166107.996	88817.70658
0.3	96560.35872	166115.88	88848.58032
0.31	96588.4351	166123.764	88879.45406
0.32	96616.51149	166131.648	88910.32781
0.33	96644.58787	166139.532	88941.20155
0.34	96672.66426	166147.416	88972.0753
0.35	96700.74064	166155.3	89002.94904
0.36	96728.81702	166163.184	89033.82278
0.37	96756.89341	166171.068	89064.69653
0.38	96784.96979	166178.952	89095.57027
0.39	96813.04618	166186.836	89126.44402



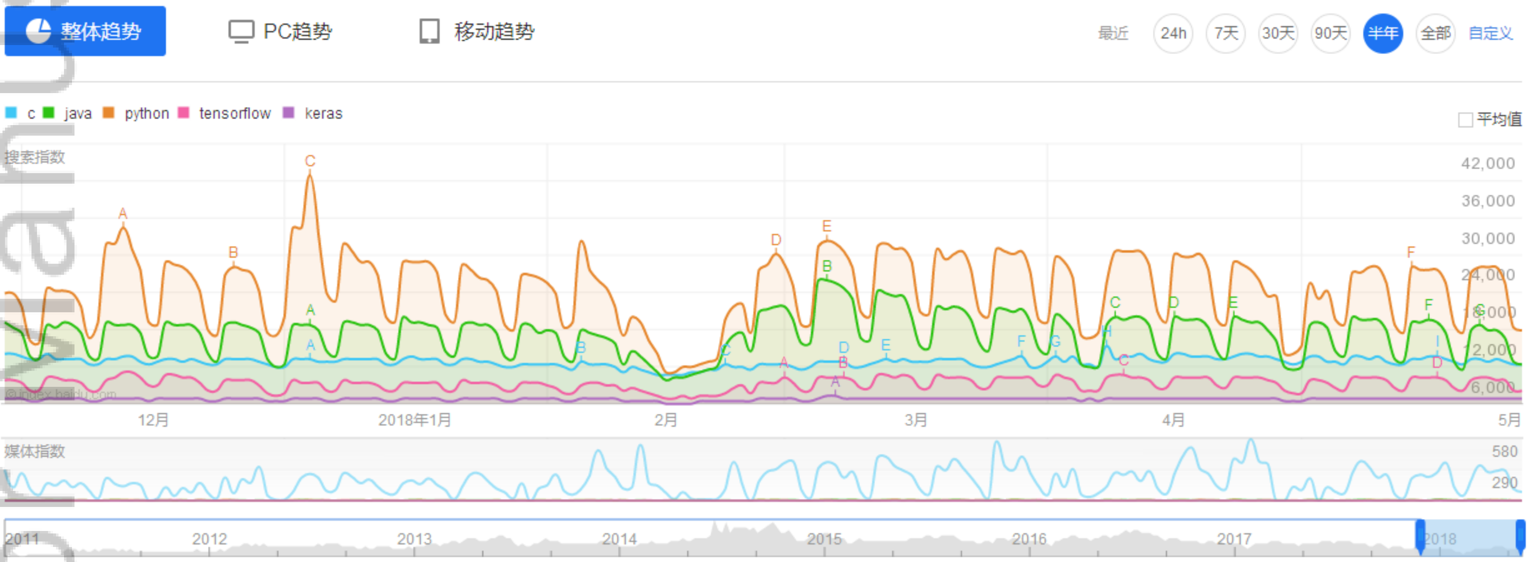
CPE_6096_Fig.3 Varying curve of the search indexes of entertainment-related words in Baidu.tif



CPE_6096_Fig.4 Framework of K-ear.tif



CPE_6096_Fig. 1 Varying curve of the search indexes of the different seasonal clothing in Baidu.tif



CPE_6096_Fig. 2 Varying curve of the search indexes of work-related words in Baidu.tif