energies

Article

Escope: An Energy Efficiency Simulator for Internet Data Centers

Jun Liu [1], Longchuan Yan [1], Chengxu Yan [2], Yeliang Qiu [2], Congfeng Jiang [2,*], Yang Li [1], Yan Li [1] and Christophe Cérin [3]

1 State Grid Co., Ltd., Information Communication Branch, Beijing 100761, China
2 School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
3 IUT de Villetaneuse, Département d'Informatique ,University of Paris 13, Sorbonne Paris Cité, LIPN/CNRS UMR 7030, F-93430 Villetaneuse, France
* Correspondence: cjiang@hdu.edu.cn; Tel.: (+86)15968100618

Abstract: Contemporary megawatt-scale data centers have emerged to meet the increasing demand for online cloud services and big data analytics. However, in such large-scale data centers, servers of different generations are installed gradually year by year, making the data center heterogeneous in computing capability and energy efficiency. Furthermore, due to different processor architectures, complex and diverse load dynamic changing, business coupling, and other reasons, operators pay great attention to processor hardware power consumption and server aggregation energy efficiency. Therefore, the simulation and analysis of the energy efficiency characteristics of data center servers under different processor architectures can help operators understand the energy efficiency characteristics of data centers and make the optimal task scheduling strategy. This is very beneficial for improving the energy efficiency of the production system and the entire data center. The Escope simulator designed in this study can simulate the online quantity (placement strategy) of different types of servers in the data center and the optimal operating range of the servers. The purpose of this is to analyze the energy efficiency characteristics of all servers in the data center and provide data center operators with the energy efficiency and energy proportionality characteristics of different servers, improve server utilization, and perform reasonable scheduling. Through the simulation experiment of Escope, it can be proved that running the server at the highest energy efficiency point or running the server under full load cannot improve the energy efficiency of the entire data center. The simulation algorithm provided by Escope can select the optimal set of servers and their corresponding utilization. Escope can set up a variety of simulation strategies, and data center operators can simulate data center energy efficiency according to their own needs. Escope can also calculate the power cost savings of introducing new servers in the data center, which provides an essential reference for operators to purchase servers and design data centers.

Keywords: data center; power consumption; energy proportionality; energy efficiency

Citation: Liu, J.; Yan, L.; Yan, C.; Qiu, Y.; Jiang, C.; Li, Y.; Li, Y.; Cérin, C. Escope: An Energy Efficiency Simulator for Internet Data Centers. *Energies* **2023**, *16*, 3187. https://doi.org/10.3390/en16073187

Academic Editors: Dimitrios I. Doukas and Antonios Marinopoulos

Received: 8 February 2023
Revised: 23 March 2023
Accepted: 29 March 2023
Published: 31 March 2023

1. Introduction

1.1. Energy Efficiency Indicators for Data Centers

With increasing demand for cloud computing from large-scale internet applications, enterprise-level essential services, and the construction of the digital economy, the scale and number of data centers have achieved unprecedented development. At the same time, the rapid growth in size and quantity has brought about many problems for data center operators, such as high energy consumption, huge cost, and severe pollution [1,2]. In sharp contrast to the sizable energy consumption of data centers, the resource utilization of most data centers is much lower [3–5], and the utilization of the majority of servers is generally less than 30%. Increasing server utilization can improve a data center's energy efficiency. Placing online services and batch jobs on the same cluster is an effective way to enhance the utilization of modern data center clusters [6,7]. However, it is not easy to deploy different types of applications to increase server utilization without affecting application performance. Their coexistence is a dilemma because it attempts to improve resource utilization while the performance of online services declines as resource utilization increases. Moreover, oversupply of peak power usage, fluctuating data traffic, and multi-level power transmission infrastructure in large data centers can lead to serious power budget fragmentation and inefficient power utilization [8].

Therefore, improving the energy efficiency of data centers has become one of the main goals of data center construction and operation. Servers are the most essential infrastructure of data centers, and their energy efficiency (EE) and energy proportionality (EP) have become hot research topics in academia and industry. If the energy efficiency of data center servers can be improved, the overall data center energy efficiency will also be effectively improved.

Data center energy efficiency indicators play a very important role in data center construction and operation management. Nowadays, there are a variety of evaluation indicators for data center energy efficiency in the industry. Power efficiency (power usage effectiveness, PUE) is one of the most important indicators to evaluate data center energy efficiency. PUE represents the ratio of total data center energy consumption to IT equipment energy consumption. The total energy consumption of a data center includes IT equipment energy consumption, cooling energy consumption, and lighting energy consumption, among other things. The energy consumption of IT equipment includes server energy consumption, network equipment energy consumption, and storage equipment energy consumption.

This study is mainly based on the energy efficiency characteristics of the servers in the data center. Improving the energy efficiency of all servers in the data center is also the top priority for maintaining the high energy efficiency of the data center. The energy efficiency of the data center is defined as Formula (1), where T_{total} represents the total number of tasks that the data center can handle, and P_{total} represents the total power consumption of all running servers in the data center:

$$\text{Date Center Energy Efficiency} = T_{total}/P_{total}. \tag{1}$$

The energy efficiency of this data center is expressed as the number of load tasks per watt and increasing the size of this metric means that when the number of tasks in the data center is fixed, the data center PUE is reduced by reducing the power consumption of IT devices.

1.2. Energy Efficiency and Energy Proportionality of Servers

The energy efficiency of data centers has become one of the main issues to be considered in the construction and management of data centers [9–14]. The most important component of a data center is servers, as shown in Formula (2). The efficiency of a server is defined as the ratio of server performance to power:

$$\text{Server Energy Efficiency} = \text{Performance/Power}. \tag{2}$$

The higher the energy efficiency of a data center or server, the more tasks that can be completed per watt of electricity. The energy efficiency indicator on a single server is often used to describe the ratio of server performance to power consumption under a certain utilization rate. Jiang [15] et al. found that the current peak energy efficiency of servers has shifted from 100% utilization level to 70–80% utilization level, which shows that it is more energy-saving to keep each server running within its energy efficiency peak range than at 100% utilization level. Figure 1 shows the energy efficiency of a server released in 2019 under different utilization rates. It can be seen that under different utilization rates, servers exhibit different energy efficiency characteristics.

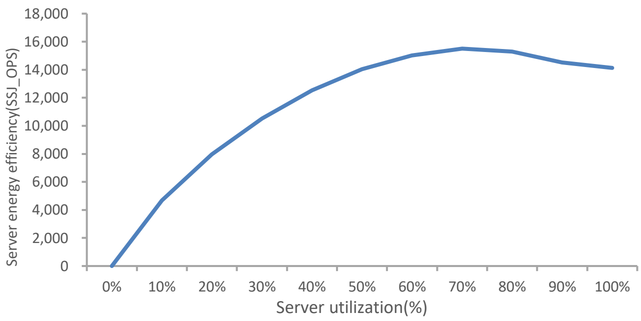

Figure 1. The relationship between a server utilization rate and energy efficiency released in 2019.

In a data center, the utilization rate of the servers is dynamic. Simply comparing the energy efficiency value of a server under a certain utilization does not necessarily mean that the server is energy saving in all cases. For example, if a server has high energy efficiency at 70% utilization, it does not necessarily mean that this server has high energy efficiency even under low utilization. In a data center with low utilization, many servers are idle. The use of such servers does not automatically improve the energy efficiency of the data center. In these cases, another indicator is needed to determine the overall energy efficiency of the servers. This indicator is called energy proportionality (EP). Energy proportionality refers to the change in server energy consumption with utilization rate. It was proposed by Rysckbosch in 2007 [16], and its Formula (3) is as follows:

$$\mathrm{EP} = 1 - (Area_{real} - Area_{ideal})/Area_{ideal}. \quad (3)$$

The server with the most ideal energy consumption curve has the following characteristics: Assuming that the power consumption of the ideal server is 100 W under full load (100% utilization), the power consumption should be reduced equally with the reduction in the load, and the power consumption at 80% load should be 80 W. When it is completely idle, it should be 0 W. At this time, the energy proportional property of this server is one, which is the ideal energy proportional characteristic. However, in actual scenarios, the server still needs power consumption when it is completely idle, so the energy consumption of the real server is not proportional to the utilization rate. As shown in Figure 2, the $Area_{real}$ in Formula (3) represents the area between the real energy proportional curve of the server and the abscissa; ideally, $Area_{ideal}$ represents the area between the energy proportional curve and the abscissa. The value of EP ranges from 0 to 2.

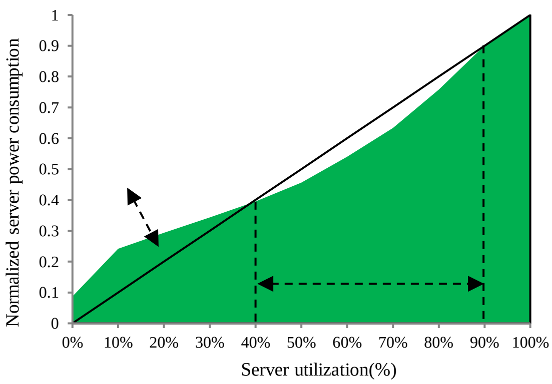

Figure 2. Energy efficiency curve of the server with an EP value of 0.98.

Wong explained the meaning of EP on paper [17]: the EP value represents the change in power consumption of a server with utilization; the energy proportionality of data center servers has significantly improved in the past decade. Energy proportionality has a great impact on the power consumption of servers. In their experiments, Barroso et al. [18]

found that servers with energy proportionality closer to one can save more power, so they recommend that optimizing energy proportionality should be the primary goal of server design. Ryckbosch et al. [16] studied the situation in which servers with higher EP values can save power consumption. Under 10–50% utilization, servers with higher EP values can ideally save 34% of energy consumption. Therefore, energy proportionality is of great significance to server research and design. Improving the energy proportionality of the server can greatly reduce the power consumption of the server. As the server energy proportionality (EP) increases, data centers have an increasing demand for peak energy efficiency-aware scheduling.

1.3. Data Center Simulation Tool

When designing a new data center, energy efficiency is one of the most concerning issues for data center operators. In large-scale data centers, servers of different generations are gradually installed year by year, which makes the data center heterogeneous in terms of computing power and energy efficiency. Additionally, in existing data centers, internal server distribution and infrastructure layouts that have been in operation for some time may undergo significant changes, as will their original energy efficiency. Therefore, the extensive energy efficiency of the data center depends to a large extent on the working range and utilization level of each server. Although the current PUE (power usage efficiency) of large data centers is already very low, if the server is not operating at full load working within the maximum energy efficiency range, then the low value of PUE is meaningless. Therefore, according to the workload characteristics, it is more energy-efficient to let different servers operate within their most energy-efficient working ranges than to let all servers operate at the same utilization level. This will greatly improve the energy efficiency of the data center.

In order to design a green data center with ideal PUE, many researchers have designed various data center simulation tools. The simulated objects include data center servers, network traffic, task placement, and heat maps, among others. The focus is on the power consumption, communication, and application response time within the data center The main problem of these simulation tools is poor stability and scalability.

Since workloads may change dynamically throughout the day, it is a challenge to adaptively select a specific subset of servers to perform tasks while powering off the power of the remaining servers so that the data center can operate with higher energy efficiency as much as possible. Due to the paucity of energy efficiency simulators and incomplete understanding of the overall energy efficiency of the data center, it is difficult for data center operators to model the energy efficiency of the entire data center. Therefore, we designed and implemented a data center energy efficiency simulator, namely Escope, which can simulate and evaluate the energy efficiency of a data center.

The difference between the data center energy efficiency simulator designed in this study and the above data center simulation is that it can simulate data center energy efficiency. According to the different workload conditions and server configuration of the data center, Escope explicitly models and visualizes the energy usage of the data center and uses the energy efficiency characteristics of the servers in the data center to formulate work server placement strategies to keep the servers at their most energy-efficient scope of work. Through the simulation of data center energy efficiency, data center throughput, server power consumption, power distribution, and other contents, the optimization algorithm is used to simulate and evaluate the data center energy efficiency as a means to help data center operators grasp the energy efficiency characteristics of each server and better perform task scheduling.

Inspired by this, we have designed a new data center energy efficiency simulator called Escope. We make the following contributions:

(1) Escope: This paper proposes the idea of simulating data center energy efficiency, which is different from other simulators. Escope can clarify the energy usage of data

centers according to different workload conditions and server configurations and model and visualize them;

(2) Flexibility: Escope allows for a variety of simulation strategies, and data center operators can simulate energy efficiency according to their own needs;

(3) Depth: Through Escope's simulation experiments, it has been shown that improving energy efficiency in data centers is not simply a matter of running servers at their highest energy efficiency point or running servers at full load. Escope's simulation algorithm can select the optimal set of servers and their corresponding utilization levels to achieve optimal energy efficiency.

The organizational structure of this paper mainly consists of the following parts: Section 1 explains the concepts related to data center energy efficiency. Section 2 introduces the current relevant research content, while Section 3 expands on the key designs and algorithms of Escope. Section 4 experiments with the Escope simulator and evaluates its role. Finally, the discussion and summary are presented.

2. Related Studies

In order to design a green data center with ideal PUE, many researchers have designed various data center simulation tools. Table 1 shows a comparative analysis of the different simulation tools. Before 2009, distributed system simulators were less frequently used in cloud computing environments, so Buyya et al. [19] and Calheiros et al. [20,21] proposed CloudSim. This simulation software can produce seamless modeling and simulation of cloud computing and their upper-layer application characteristics. CloudSim supports simulation of cloud computing infrastructure and management services, so users can use CloudSim to study specific system problems. CloudSim also introduced the simulation of virtualized data centers and used NetworkCloudSim to expand the functions to better support the simulation of internal communication in the data center [22]. However, the scalability of CloudSim is poor, and experiments prove that the CloudSim simulator will encounter various failures during the submission of the job.

Table 1. Comparison of different simulation tools.

	Characteristics	Advantages	Disadvantages
CloudSim	Cloud computing environment simulation	Open source, user-friendly	Poor scalability
CloudAnalyst	Performance evaluation, cost-benefit analysis	Visualization	Not integrated with the latest version of CloudSim functionality
DartCSim+	Resource allocation, load balancing	Simulates large-scale cloud environments	Steep learning curve
DVFStoCloudSim	Energy-aware simulation	Integrated with DVFS technology	Considers only CPU power consumption
WorkflowSim	Simulates workflow processes	Supports multiple workflow models	No dynamic resource scheduling algorithm

In order to reduce or even eliminate the shortcomings and faults in CloudSim, many researchers [23–30] improved the CloudSim simulator. For example, Li et al. [25] designed the simulator DartCSim+, which supported power-aware network simulation. In order to solve transmission failures caused by migration or network failures, DartCSim+ uses a resubmission mechanism based on packet transmission. Bux et al. [26] solved the inhomogeneity problem of CloudSim. On the basis of CloudSim, they added a process of modeling instability in the cloud environment. It could simulate dynamic changes in runtime performance and sudden changes during task execution issued by the failure. In addition, Guérout and Monteil et al. [27] and others added a new patch of DVFS to CloudSim, so that CloudSim could use DVFS to perform energy-aware simulation experiments. Chen and Deelman et al. [28] et al. pointed out that ignoring the system failure and overhead in the simulation workflow would have a significant impact on the simulation experiment results, so they proposed WorkflowSim, which can be used to simulate the workflow in a distributed environment. Wickremasinghe [29,30], also from the Buyya team, proposed the visualization simulator CloudAnalyst, based on CloudSim. The main purpose of the simulator was to achieve optimal scheduling of the data center under the current configuration conditions. CloudAnalyst was designed directly based on CloudSim and expanded some of CloudSim's functions. It could be used to learn the behavior of large-scale internet applications in the cloud computing environment and quickly conduct simulation experiments. Kecskemeti [31] introduced a unified model of resource sharing and a hierarchical energy monitoring framework, thus solving the scalability problem of CloudSim.

Based on CloudSim, Tian et al. [32] developed the lightweight visual cloud computing simulator CloudSched, which could support simulation modeling of large-scale cloud computing applications. Using this simulator, users can customize their information, data center information (number and location, etc.), resources, and other information and simulate basic indicators such as data center response time and processing requests. However, the simulator cannot simulate the amount of power consumption of the data center. The MDCSim simulator [33] simulates the power consumption of the data center and can model the characteristics of various devices (servers, switches, etc.) in the data center. MDCSim can avoid building and processing similar simulation objects one by one, so the required simulation time is significantly shortened, and the scalability is significantly improved. CloudSim and MDCSim are event-based simulators. Their simulation accuracy is insufficient. The MDCSim simulator is a commercial product, and its working principle cannot be understood due to the lack of public source code. In order to improve the simulation accuracy, Kliazovich [34] proposed a new simulator, GreenCloud, which is a packet-level cloud data center simulator designed to evaluate the energy consumption cost of data center operation. This simulator is an extension of the network simulator NS2 [35]. It mainly focuses on evaluating the power consumption of cloud communications and provides a fine-grained power modeling and simulation tool for cloud data centers. Its key advantage is that it fully supports the TCP/IP protocol model. For fine-grained simulation of data center power consumption, DCWorms [36] provides simulations of data center energy consumption, including energy consumption of cooling and ventilation systems and energy consumption modeling of CPU, memory, and network in servers.

In order to support elastic cloud infrastructure simulation, Sriram [37] proposed the SPECI simulation tool. It allows simulation of the performance and behavior of data centers and simulates the functions and code of large data centers according to input size and the middleware design strategy. SPECI consists of two packages: one for building data center layouts and topologies and the other for executing experimental components, so it has good scalability. Unlike the simulators introduced above, which are software-based simulators, OpenCirrus [38,39] is an open cloud computing simulator based on software and hardware, designed to support server design and management research in data centers. The simulator has three main goals: to promote system-level research on cloud computing; to encourage new cloud computing applications and application-level research; and to provide experimental data sets to supply open API for cloud computing development.

The difference between the energy efficiency simulator designed in this research and the above data center simulation is that Escope mainly simulates the energy efficiency of the data center. By simulating data center energy efficiency, data center throughput, server power consumption, and power distribution, optimized algorithms are used to simulate and evaluate data center energy efficiency. It can also help data center operators master the energy efficiency characteristics of each server, as a means to better perform task scheduling.

3. Data Center Energy Efficiency Simulator—Escope Design

Experiments have shown that running servers in the best working range can improve the energy efficiency of servers. Still, hundreds of thousands of servers are deployed in large data centers, hundreds of which have different configurations and models. Due to different hardware configurations, the energy efficiency characteristics of servers are also different [40,41]. Therefore, data center operators need to understand the energy efficiency and energy ratio of each server in order to effectively improve server utilization and reasonable task scheduling with the objective of improving the energy efficiency of the data center as much as possible. In order to enable data center operators to better understand the energy efficiency characteristics of servers in the data center, we have developed the data center energy efficiency simulator Escope. We use Escope to simulate the online number of servers of different models in the data center and the optimal working range of the servers to analyze the energy efficiency characteristics of the data center servers and provide data support for data center operators. By modeling data center energy efficiency, task throughput, and power quota, optimization algorithms simulate and evaluate data center energy efficiency to help data center operators understand the energy efficiency characteristics of each server, thus improving task scheduling and data center energy efficiency.

3.1. System Function Design

The architecture of Escope is shown in Figure 3. It includes components such as crawler, simulator, database, selector, and web interface.

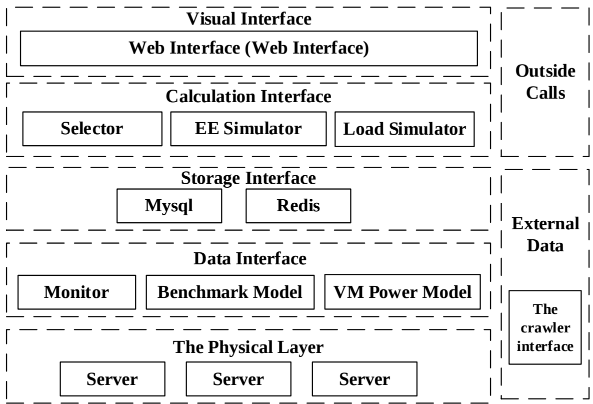

Figure 3. Escope architecture.

Escope provides the following functions:

(1) Escope integrates the monitoring solution for large-scale distributed clusters in the open-source framework Zabbix, which can collect runtime data of all servers in the cluster, including temperature, CPU utilization, memory utilization, server power consumption, memory power consumption, and power management unit (PMU) information. Monitoring items can be added at any time according to the needs of users;

(2) Escope provides a load model, which integrates a variety of commonly used loads in the industry. Users only need to enter information such as memory utilization, disk I/O, and network I/O, and Escope will automatically complete the test on the designated physical machine to obtain data information. The generated data will support the data center energy efficiency simulation;

(3) For cloud computing operators, Escope provides a VM power consumption model, which can automatically estimate the VM power consumption based on the system information collected on the server;

(4) In addition to automatically testing the energy efficiency of the server according to the load model, Escope can also automatically obtain the server configuration and energy efficiency information (such as SPECpower) disclosed by each website to establish a server energy efficiency information database;

(5) Users can make simulation policies and scenarios, such as setting the utilization rate of all servers to no more than 30% or setting the utilization range for servers to simulate the load of different business scenarios;
(6) Users can add new servers and server prices to the constructed data center. After adding new machines, Escope can simulate the energy efficiency operation of the newly constructed data center and calculate the cost–benefit based on the electricity savings;
(7) Escope provides a web visual interface, so users can input various parameters and display the simulation results. For example, it will classify and display the simulation results according to server parameters (CPU model, memory size, release year, etc.) to deeply analyze the energy efficiency distribution of data center servers.

3.2. Interface Design

The functions of each interface in Escope are as follows:

(1) Data interface: This interface is mainly used to test and collect energy efficiency information of data center servers, including three components: monitoring, energy efficiency testing, and VM power consumption model. The monitoring component is used to collect information about the server. The energy efficiency testing component integrates a variety of loads to perform energy efficiency testing on the server. The VM power consumption model is used to estimate the power consumption of the VM based on the monitoring information of the server;
(2) External data acquisition interface: This interface is responsible for acquiring server energy efficiency information from the website that publishes server energy efficiency information, sorting the data and storing it in the Escope database; hence, even without real servers, Escope can still simulate data center energy efficiency;
(3) Storage interface: The function of the storage interface is to store and classify server information obtained from various methods. Escope uses MySQL for persistent storage and Redis for data caching, thus speeding up simulation efficiency while ensuring data integrity;
(4) Calculation interface: This interface includes three components, namely, the selector, the energy efficiency simulator, and the load generator. The selector can select qualified server information and quantity from the database according to the parameters provided by the user to construct the data center to be simulated. The load generator can estimate the total throughput of the data center based on the typical load situation of the data center and then input it into the simulator for simulation. The simulator contains an energy efficiency simulation algorithm, which executes the simulation algorithm according to the simulation strategy parameters input by the user. Then, it generates a report after the simulation ends and displays the simulation result on the Web interface;
(5) Visual interface: The visual interface has two functions. The first is to receive simulation parameters provided by users, such as simulation strategies and server combination strategies. The second is to display Escope simulation results for users to analyze;
(6) External call: Escope provides an external call interface, and other systems in the data center can call the simulation results generated by Escope through RPC (remote procedure call).

The detailed execution process of Escope is shown in Figure 4. (The server energy efficiency test process is omitted.)

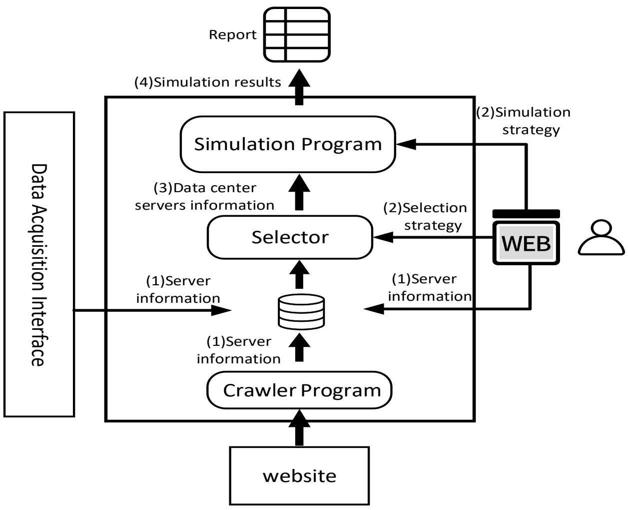

Figure 4. Escope Simulation Process.

(1) First, users collect information about the server's energy efficiency, either by crawling from some web sites (SPECpower, for example) or by entering custom server information through a web interface;
(2) Then, users enter the selection parameters, the selector selects the server according to the parameters and builds the data center to be simulated. The parameters can be server type, server release year, server quantity, and CPU type;
(3) Next, the user inputs a simulation strategy, which can be to limit the total power consumption of the data center or the throughput that the data center needs to achieve. The goal of the simulator is always to select the server and its utilization rate under the existing simulation strategy to maximize the energy efficiency of the data center, thereby maximizing the throughput and energy efficiency of the data center;
(4) Finally, when the simulation is completed, the simulator will send the results to the Web server in JSON format and generate a corresponding result report. The web interface will display different types of simulation result data and charts according to the simulation result, which is convenient for users to view and analyze the data.

3.3. Energy Efficiency Simulation Algorithm for the Data Center

The Escope energy efficiency simulation algorithm needs to solve a combinatorial optimization problem. Assuming that the data center has x servers with different configurations, the number of servers in each configuration is n, and there are N servers in total. Each server can run underutilization j ($j = 1, 2, 3...10$, representing 10% to 100% utilization).

$$\begin{gathered} \max \textstyle\sum_{i=1}^{i=x} \sum_{j=1}^{j=10} q_{ij} n_i c_i, \\ \text{s.t.} \textstyle\sum_{i=1}^{i=x} \sum_{j=1}^{j=10} p_{ij} n_i c_i \leq P, \\ \text{s.t.} \ \textstyle\sum_{j=1}^{j=10} n_i \leq n. \end{gathered} \tag{4}$$

The power consumption of server i under the utilization rate of j is p_{ij}; the throughput is q_{ij}, and the value of c_j is either 0 or 1, which represents whether to turn on the server. Formula (4) expresses the maximum throughput of the data center under the power limit, and Algorithm 1 shows the throughput maximization algorithm under the power limit.

Algorithm 1: Maximizing throughput simulation algorithm

Input: Data center server set N, Total power P

Output: Optimal solution bestvalue

```
for k in [1 ,P] :
  for i from [1, N + 1] :
     if n_i is failure :
          bestvalue[i, j] = bestvalue[i − 1, j]
     end if
     for j from 1 to 10 :
        if k < p_ij:
            bestvalue[i, k] = bestvalue[i − 1, k]
        else if bestvalue[i − 1, k] ≥ q_ij and bestvalue[i − 1, k] >
bestvalue[i, k]:
            bestvalue[i, j] = bestvalue[i − 1, k]
        else :
            bestvalue[i, k] = q_ij + bestvalue[i − 1, k − p_ij]
        end if
     end for
  end for
bestValue = bestvalue[N + 1, k + 1]
```

Algorithm 1 is similar to the multiple knapsack algorithm. The two-dimensional array bestvalue [i, j] represents the maximum throughput of the first i server under power consumption j, and bestValue represents the maximum throughput of the data center under power consumption limit P. In this study, the power consumption limit P of the data center is regarded as the backpack capacity, the power consumption of the server under the utilization rate j as the weight, and the number of tasks that the server can handle under the utilization rate j as the value. The function of the algorithm is to aggregate the appropriate server and utilization and maximize the target value under restricted conditions. As shown in Formula (5), Escope can also minimize the energy consumption of the data center while limiting the total number of tasks in the data center.

$$\begin{gathered} \min \textstyle\sum_{i=1,j=1}^{i=x,j=10} p_{ij} n_i c_i, \\ s.t. \textstyle\sum_{i=1,j=1}^{i=x,j=10} q_{ij} n_i c_i \geq T, \\ s.t.\ \textstyle\sum_{j=1}^{j=10} n_i \leq n. \end{gathered} \tag{5}$$

The data center power minimization algorithm is shown in Algorithm 2. The bestValue in the algorithm represents the minimum power consumption required by the data center when the data center throughput is T.

Algorithm 2: Minimizing total power consumption simulation algorithm

Input: Data center server set N,Total throughput T

Output: Optimal solution $bestValue$

```
P = 0
for i in (1, N + 1) :
    P = P + p_{i,10}        // maximum power consumption of data center servers
for k in (1, P):
  for i in (1, N) :
    if n_i is failure:
        bestvalue[i, j] = bestvalue[i − 1, j]
      for j from 1 to 10:
        if k < p_ij:
            bestvalue[i， k] = bestvalue[i − 1. k]
        else if bestvalue[i − 1, k] ≥ q_ij and bestvalue[i − 1, k] > bestvalue[i, k]:
            bestvalue[i, j] = bestvalues[i − 1, k]
        else:
            bestvalue[i, k] = q_ij + bestvalue[i − 1, k − p_ij]
        end if
      end for
  end for
while(bestvalue[i, k] > T) :
    k—
end while
bestValue = bestvalues[N + 1, k + 1]
```

Based on Algorithms 1 and 2, users can calculate the optimal value under the data center's total power consumption or throughput limit. Assuming that the set of servers selected by Algorithm 1 and Algorithm 2 is S, Algorithm 3 can output the optimal server combination and the specific utilization rate of these servers according to bestvalue [*i*, *j*]. The two-dimensional array utl [*i*, *j*] represents the utilization rate selected by the server n_i under the limit j, and the utilization rate ranges from 10% to 100%. When bestvalue [*i*, *L*] is greater than bestValue [*I* − 1, *L*], server n_i will join the server set S, and the utilization rate of server n_i will be recorded at this time. When the simulated data center is too large, the two-dimensional array bestvalue [*i*, *j*] may cause memory overflow. When j exceeds the threshold (the size of the threshold is related to the size of the JVM), Escope will divide a large knapsack problem into multiple small knapsack problems for simulation calculation, and the segmentation accuracy will be lost (less than 1%). Although the accuracy is reduced, it ensures that Escope can simulate data centers of any size. Users can freely adjust the threshold according to the configuration of the computer running Escope, thus ensuring the scalability of Escope.

Algorithm 3: Server Selection Result output

Input: $bestvalues[][]$

L $=$ Total throughput T || Total power P

Data center server set N

Output: Selected server set S

$Initialize\ array\ utl[n+1]\ [L+1]$

$\boldsymbol{for}\ i\ in\ (n+1,1)$:

  $\boldsymbol{if}\ bestvalues[i][L] > bestvalues[i-1][L]$:

    $S.add(n_{i-1})$

  $\boldsymbol{end\ if}$

  $\boldsymbol{if}\ L\ is\ Power$:

    $L = L - p_{i,utl[i-1][L]}$

  $\boldsymbol{end\ if}$

  $\boldsymbol{if}\ L\ is\ throughput$:

    $L = L - q_{i,utl[i-1][L]}$

  $\boldsymbol{end\ if}$

  $\boldsymbol{if}\ L == 0$:

    $Break$

  $\boldsymbol{end\ if}$

$\boldsymbol{end\ for}$

Although the above algorithm can obtain the optimal solution of the combinatorial optimization problem, its time complexity reaches $O(n^3)$, which requires more time to simulate a large data center with a large number of servers. Therefore, this research also integrates a simulated annealing algorithm (simulated annealing, SA) in Escope. The simulated annealing algorithm is a method of seeking approximate solution optimization problems based on a Monte Carlo design. The simulated annealing algorithm is essentially a greedy algorithm. Because it adds random factors in searching for the optimal solution, it has a certain probability to accept the sub-optimal solution, which may jump out of the local optimal solution and reach the global optimal solution.

As shown in Figure 5, assuming that the minimum point C is the optimal solution, the simulated annealing algorithm will continue to move to the right with a certain probability after searching for the local optimal solution B. By moving to the right, there is a certain probability that B and C can be skipped. Therefore, the local minimum B is jumped out, and the optimal value C is reached. The probability of accepting the sub-optimal solution adopts the metropolis criterion. As shown in Equation (6), the probability that the particle tends to balance at temperature T is $\exp(-\Delta E/(kT))$, where E is the internal value at temperature T, ΔE is the variable, and K is the Boltzmann constant.

$$\begin{cases} 1 & E(x_{new} < E_{old}) \\ \exp\left(-\dfrac{E(x_{new}) < E(x_{old})}{T}\right) & E(x_{new} \geq E(x_{old})) \end{cases} \tag{6}$$

The simulated annealing algorithm is shown in Algorithm 4. When it is used to solve the combinatorial optimization problem in this section, the internal energy E can be assumed as the data center throughput, and the temperature T can be simulated as the control parameter t to perform the simulated annealing algorithm. First, starting from the initial i and the initial value of the control parameter t, the current solution is repeated to generate a new solution (delete or add a new server); calculate the objective function difference (compared with the previous data center throughput); accept or discard iteration of the solution (whether the restriction condition exceeds the threshold); and gradually attenuate t (recalculate the probability of accepting the sub-optimal solution). Finally, the server combination generated at the end of the algorithm is the approximate optimal solution.

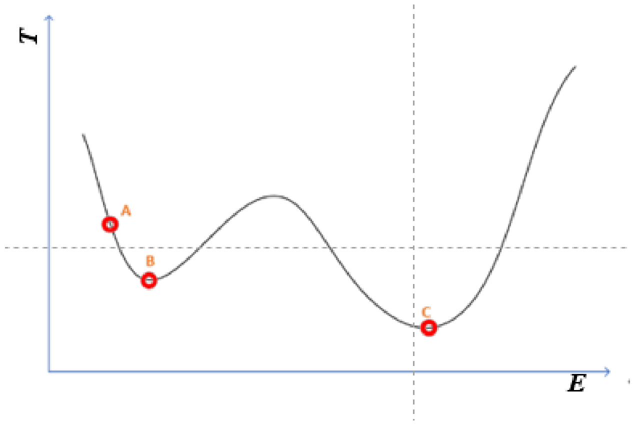

Figure 5. Simulated annealing algorithm to find the optimal value (the lowest point).

Algorithm 4. Simulated Annealing algorithm.

Input: Data center server set N, Total throughput T,
Initial temperature I_t, Annealing rate a_f, number of balances b
Number of iterations $iter$
Output: Selected server set S

$\boldsymbol{for}\ i\ in\ (0, iter):$
  $\boldsymbol{for}\ j\ in\ (0, b):$
    $nowValue = calcuValue()$
    $server = random()$
    $S = put(server)$
  $\boldsymbol{if}\ calcuWeight() > P:$
    $delete(server)$
    $continue$
  $\boldsymbol{end\ if}$
  $\boldsymbol{if}\ calcuValue() >= nowValue:$
    $continue$
  $\boldsymbol{else}:$
    $Math.random() < Math.exp(\frac{cacluValue - nowValue}{I_t})$
  $\boldsymbol{end\ if}$
  $\boldsymbol{end\ for}$
  $I_t = I_t * a_f$
$\boldsymbol{end\ for}$

The annealing rate in the simulated annealing algorithm has an undeniable impact on the efficiency of the algorithm. Although the temperature drops too fast to reach stability quickly, it will reduce the probability of obtaining the optimal solution. If the temperature drops too slowly, the algorithm will take too much time. In this study, the experimental annealing rate is 0.95. The more you set the number of balances, the fewer iterations you need, but the time for a single iteration becomes longer. The setting of the initial temperature will affect the search range of the solution. The higher the temperature, the higher the quality of the final solution, but the algorithm will take longer. Compared with the knapsack algorithm, the simulated annealing algorithm can obtain sub-optimal solutions, but the time complexity is reduced to O $(2x(\log(n))^2)$. We use two different simulation algorithms in Escope to minimize power consumption for 10,000, 100,000, and 1 million servers. The input is the number of tasks that need to be processed. Algorithm 2 is labeled as Algorithm 1, and Algorithm 4 Labeled as Algorithm 2. The experimental results are shown in Table 2.

Table 2. Simulation time comparison of different algorithms.

	Number of Servers (10^4)	**Number of Servers (10^5)**	**Number of Servers (10^5)**
Algorithm #1 completion time	54 s	228 s	1498 s
Algorithm #1 accuracy	100%	100%	99.4%
Algorithm #2 completion time	8 s	19 s	488 s
Algorithm #2 accuracy	99.1%	97.6%	94.1%

Algorithm #1 has a smaller accuracy in the case of one million units. The reason is that there are too many servers to be simulated, which requires decomposition and processing, resulting in a decrease in simulation accuracy. On the other hand, Algorithm #2 has a faster calculation time. In the case of a million scale, the speed is 68% higher than Algorithm #1, but it also loses 6% accuracy.

Data center operators can choose different simulation algorithms according to their needs. If a fast simulation is required and the optimal value is not required, then the simulated annealing algorithm can be selected. If the optimal solution is needed, the multiple knapsack algorithm can be used for simulation.

4. Experiment and Analysis

In order to verify the effectiveness of energy efficiency simulator Escope, four kinds of simulation data centers were established in this section. The data center server was constructed with the server energy efficiency information released by SPECpower in 2017–2019; 135 types of servers were selected, each type of the server was set to 50, and the total number of servers in a data center was 6750. We set two different simulation goals. One goal was to calculate the number of servers online when the data center throughput was maximized and the utilization group located under the power consumption limit of the data center. The second goal was to calculate the online number of servers that minimize the power consumption of the data center and the utilization rate when the task load of the data center was determined. Escope would simulate the optimal solution for the highest energy efficiency in the data center by simulating which servers were turned on and at which utilization rate among the 6750 servers. The overview of the four different types of data centers is as follows:

(1) Data center #1: the server can choose to run at any utilization;
(2) Data center #2: the server is always running at the highest energy efficiency;
(3) Data center #3: server utilization is less than 30%;
(4) Data center #4: the server runs at 100% utilization.

In the simulation of this article, data center #1 has no restrictions on server utilization, and all servers can choose to operate at any utilization. Intuitively, as long as each server runs on the highest EE, the data center can achieve the highest energy efficiency, so we set up data center #2 to verify whether this idea is correct. The server portfolio simulated by data center #2 will all run at its highest EE utilization. Data center #3 simulates a typical data center situation where the server usage rate is less than 30%. This situation wastes a lot of resources but is not completely useless. The servers selected in data center #4 will run at 100% utilization, and the purpose of the setting is to verify whether running the server at 100% utilization is an optimal policy. No matter which simulation strategy is used, Escope will select the best combination of servers under the current situation to maximize the energy efficiency of the data center while satisfying the current strategy.

4.1. Simulation of Maximum Throughput in Data Center with Limited Power Consumption

Constrained by data center power infrastructure and cooling conditions, data center servers must operate under strict power restrictions. When the power of all the servers on the rack exceeds the rated upper limit, the servers will power out, which will affect the stable operation of the data center. Therefore, the limited power consumption poses problems for data center operators: When the total power of the data center is limited, which types of servers should data center operators choose, and how many servers should run in each model? Furthermore, at what utilization range should these servers keep maximizing the throughput of the data center?

In this section, several simulation experiments will be conducted on the four aforementioned data centers to explore the energy efficiency operating range of different types of servers in data centers under power constraints. The upper limit of the total power of each data center in the experiment is 100 KW to 1000 KW, and the increased step length is 10 KW. The experimental results are shown in Figures 6 and 7.

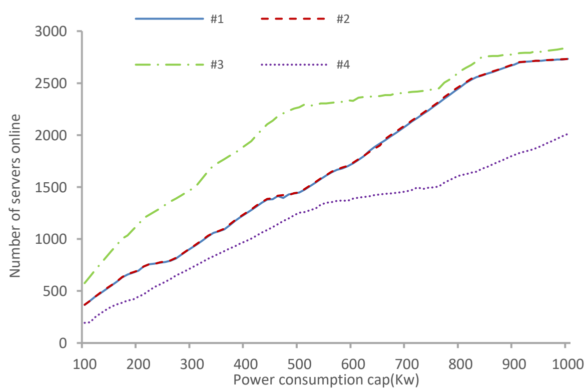

Figure 6. Number of servers online for different data centers.

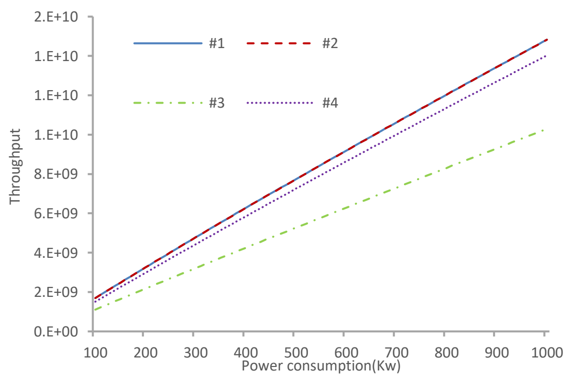

Figure 7. Throughput of different data centers.

The simulation results of data center #1 and data center #2 are similar, but the throughput of data center #1 is always greater than that of data center #2. The reason for this result is that although the simulation strategy of data center #2 always selects the server utilization at the peak EE, in some cases, this choice does not maximize the total throughput. For example, when the available power consumption is 200 W, the selected server running at 100% utilization needs 200 W, and the server running at EE peak (assuming 80%) needs 180 W. Therefore, #1 will select 100% utilization, while # 2 will choose 80% utilization, so the remaining 20 W will be wasted. Obviously, since data center #1 can choose any utilization rate and the simulated optimal combination of servers can maximize throughput, the throughput of data center #1 is always the largest.

The throughput of data center # 3 is much lower than that of other data centers, indicating that traditional data centers waste resources when server utilization is less than 30%, thus reducing the throughput and energy efficiency of the entire data center. For data center # 4, the number of servers online is always the lowest, because server power consumption is always the highest at 100% utilization, so it is easy to reach the power limit. This also verifies that the strategy of unplanned running servers at the highest utilization does not improve the energy efficiency of the data center.

By analyzing the experimental results, the following conclusions can be drawn:

(1) Under the same power consumption limit, the number of online servers at peak energy efficiency (data center #2) is about 29.86% higher than servers at 100% utilization (data center #4), and the total number of tasks in the entire data center (ssj_ops) increased by 7.17%;
(2) The number of servers in data center #1 is slightly lower than that in data center #2, but the throughput is 2% higher than that of servers running at peak energy efficiency utilization (data center #2).
(3) The server utilization rate of the traditional data center is lower than 30% (data center #3), and the average throughput is 33% lower than that of data center #1. However, the number of online servers has increased by 25% when compared to #1. The largest number of online servers means that data center #3 has better redundancy, which can ensure that the data center provides stable services.

Judging from the energy efficiency distribution of Escope's selection of server collections, the energy efficiency of servers released in recent years has been significantly improved when compared to many years ago, and the average EE and EP have been greatly improved. Table 3 shows the average EE and EP of different data centers when the power consumption is limited to 1000 KW. Comparing the average EE values of data center #1 and data center #2, it can be seen that it is a better strategy to choose a server with a high EE value, but this is not always optimal. The largest EP is chosen in data center #2, because EP represents the variation in server power consumption with utilization. Therefore, servers with large EE may not have the largest EP, but servers with larger EPs tend to have a higher EE value when working at low utilization (10–30%).

Table 3. The average EE and EP of the four data centers with a power limit of 1000 KW.

Data Center	Average EE	Average EP	Average utilization
#1	15,331	0.86	79.75%
#2	15,148	0.87	77.95%
#3	10,571	0.92	30%
#4	14,327	0.83	100%

4.2. Simulation of Power Minimization with Data Center throughput Preserved

When the size and business of the data center are stable, the throughput of the data center remains stable. Minimizing the power consumption of the data center when the throughput is fixed is also a key issue for improving the energy efficiency of the data center. In this section, the simulation goal of all four data centers is to minimize data center power consumption while maximizing data center throughput. In the simulation, the total throughput of each data center is set from 1×10^8 to 1×10^{10}, with a step size of 1×10^8.

The simulation results are shown in Figures 8 and 9. For data center #4, the number of online servers is always the lowest. This is because running at 100% utilization on the server means that the server needs to consume the most power and can handle the most tasks. However, maximum peak processing capacity does not imply highest energy efficiency. Therefore, data center #4 has an average power consumption increase of 1.7% (about 30 KW) when compared to data center #2, with the highest energy efficiency utilization. The average energy consumption of data center #1 is reduced by 0.85% when compared to data center #2. For the traditional data center #3, due to the low resource utilization rate, the average power consumption increased by 36% (about 200 KW) when compared to the other three high utilization data centers. However, the number of online servers in data center #3 has increased by an average of 45% when compared to other data centers, which can better guarantee the quality of service.

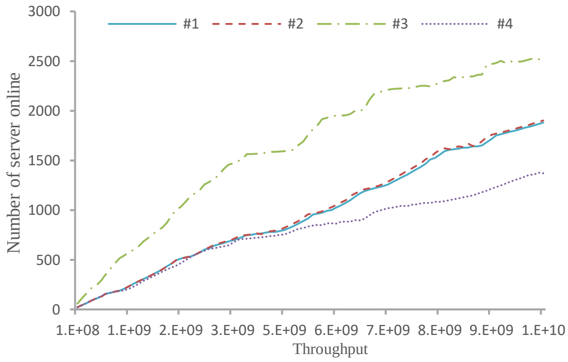

Figure 8. Number of servers online in different data centers.

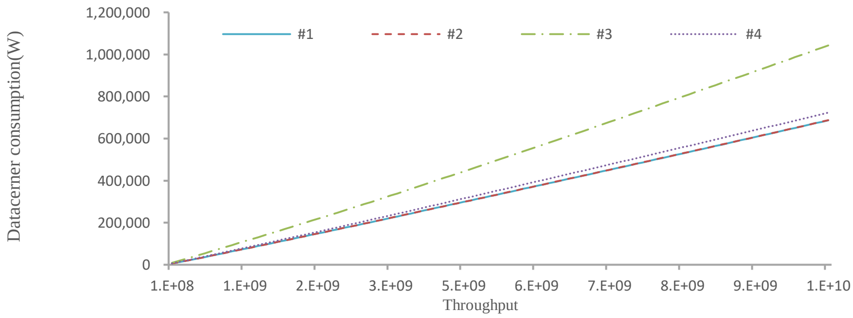

Figure 9. Total server power consumption of different data centers.

Table 4 shows the release year of the server selected by data center #1 in the simulation. It can be seen from the table that in order to achieve maximum energy efficiency, the choice of the simulator is related to the server energy efficiency characteristics but not the release year, which shows that Escope can choose the best server combination according to different strategies.

Table 4. Release year and number of servers under 1×10^{10} throughput in data center #4.

Server Name	Price (USD)	EE	Number(U)
Fusion 2288 H V5	7768	13,478	10
Fujitsu RX2540 M4	11,172	12,842	10
Fujitsu RX4770 M4	9270	12,828	10
ThinkSystem SR950	10,025	12,377	10
Sugon I820-G30	9487	12,306	10

Escope will use load generators to generate tasks that the data center needs to complete. The load generation result is shown in Figure 10, which shows the throughput of the data center in a day. The maximum task volume of the data center is 6.0×10^8, and the hourly load in the day is set to a step shape based on SPECpower modeling.

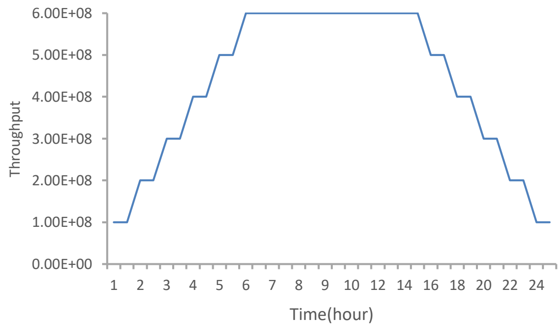

Figure 10. Throughput of the data center per day.

The simulation results are shown in Tables 5 and 6. The energy efficiency of data center #5 is 7297, and the energy efficiency of data center #6 is 13758. Based on the average electricity price of 1 CNY/kWh in China in 2019, small data center #6 compares with data center #5 by saving about CNY 1000 per day. Using the Escope simulation, it takes 3418 days to recover the cost of purchasing a new server by saving power.

Table 5. Simulation results of data center #5.

Throughput	Average Consumption (W)	Number of Severs Online	Time (hour)
1E8	11,660	36	2
2E8	23,645	47	2
3E8	36,967	71	2
4E8	52,873	140	2
5E8	70,823	162	2
6E8	91,594	210	14

Table 6. Simulation results of data center #6.

Throughput	Average Consumption (W)	Number of Severs Online	Time (hour)
1E8	6458	22	2
2E8	13,082	37	2
3E8	19,946	41	2
4E8	26,930	50	2
5E8	37,193	67	2
6E8	49,052	38	14

The larger the size of the data center, the greater the electricity consumption, the higher the electricity price, and the faster the cost recovery will be. However, this number is only a reference value for data center operators. Data center operators can simulate data center energy efficiency by configuring different server combinations so as to choose the most suitable server for their data center and reduce procurement costs.

5. Discussion

In practice, on the one hand, data center servers must run under strict power restrictions, and operators need to consider which types of servers should be selected and how many servers should be run in each model; they also need to understand what utilization range must be maintained to maximize the throughput of the data center. On the other hand, when the size of the data center and the business is stable, the throughput of the data center will remain stable. Minimizing power consumption in the data center when throughput is fixed must also be considered. Escope can simulate the energy efficiency of the data center. By entering the power consumption limit of the data center, we can simulate the maximum throughput of the data center under this power limit. Simulations can also be run to determine which servers should be started and what utilization range the servers should be run in as a means to achieve the maximum throughput. At the same time, by entering the number of tasks that the data center needs to handle, Escope can calculate the minimum power consumption required by the data center server to handle these tasks. In terms of specific algorithm selection, data center operators can choose different simulation algorithms according to their needs. If a fast simulation is required and the optimal value is not necessary, then the simulated annealing algorithm can be selected. If the optimal solution is needed, multiple simulation algorithms can be selected. The backpack algorithm is simulated.

6. Conclusions

This study starts with the reduction in data center power consumption; then, related research on data center energy efficiency and energy proportionality are introduced. Subsequently, we develop the design and experimental analysis of the energy efficiency simulator Escope, and now we can draw the following conclusions:

(1) The energy efficiency of the data center cannot be improved by running the server at the highest energy efficiency point or by running the server under full load. The simulation algorithm provided by Escope can select the optimal server set and their corresponding utilization rate;
(2) Escope can set a variety of simulation strategies, and data center operators can simulate data center energy efficiency according to their own needs;
(3) When limiting the server utilization rate to less than 30%, almost all simulation results of the server run at 30% utilization rate. This is because under the existing architecture, the energy efficiency of the server at 30% utilization rate must be higher than a utilization rate of less than 30%, so most of the selected servers run at 30% utilization rate;
(4) Escope can calculate the electricity cost saved by introducing new servers in the data center. This function provides an important reference for operators to purchase servers and design data centers.

In addition to the data centers constructed in the study, operators can also build data centers that conform to the actual situation. According to different management purposes, the data center may have different operation strategies. Escope can simulate the energy efficiency of the data center according to different strategies, help data center operators understand the energy efficiency characteristics of the data center, and require the data center task scheduler to make the best decision.

In future work, we plan to set up server performance data for more benchmark types such as hybrid web server benchmarks and memory intensive benchmarks. Further, we hope to allow users to customize the benchmark. We will add a monitoring system to Escope, allowing Escope to automatically monitor server performance. Administrators can test more benchmarks (rather than SPECpower only), and Escope can automatically generate server performance data to bring the simulation closer to reality.

In addition, due to the fact that new applications of artificial intelligence require a large amount of computing and storage resources to support their algorithms and models, these resources need to be supported and maintained in data centers, resulting in significant energy consumption. In the future, we can further use CPU and GPU to accelerate simulation calculations in data centers and improve simulation speed and accuracy. Quantum computing, on the other hand, can better handle complex problems such as optimization and derivation of artificial intelligence algorithms, thereby improving the level of intelligence in data centers. By combining these tools, different data center scenarios can be simulated, and future performance and energy costs can be predicted.

Author Contributions: Conceptualization, C.J. and C.C.; methodology, C.J. and L.Y.; software, C.Y. and Y.Q.; validation, J.L., Y.L. (Yan Li), and Y.L. (Yang Li); formal analysis, Y.L. (Yan Li) and Y.L. (Yang Li); resources, L.Y.; data curation, C.Y.; writing—original draft preparation, C.J.; writing—review and editing, L.Y.; visualization, J.L.; supervision, C.J.; project administration, L.Y.; funding acquisition, C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of State Grid Corporation of China (No.52993920002M), and the Natural Science Foundation of China (No.61972118). The funder was not involved in the collection, analysis, or interpretation of the data or manuscript.

Data Availability Statement: All data from the study are presented in the manuscript.

Acknowledgments: The authors thank all participants for their time and cooperation throughout the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cook, G. *How Dirty is Your Data: A Look at the Energy Choices that Power Cloud Computing*; Greenpeace: , Amsterdam, The Netherlands,2011.
2. Pettey, C. Gartner estimates ICT industry accounts for 2 percent of global CO2 Emissions. *Gartner Symposium/ITxpo* **2007**,San Francisco,http://www.gartner.com/us/symposiumwest.
3. Jyothi, S.A.; Curino, C.; Menache, I. *Morpheus: Towards Automated SLOs for Enterprise Clusters*; In Proceedings of USENIX OSDI 2016; USENIX, Berkeley, California, pp. 117–134.
4. Rajan, K.; Kakadia, D.; Curino, C. PerfOrator: Eloquent performance models for Resource Optimization. In Proceedings of the Seventh ACM Symposium on Cloud Computing, Santa Clara, CA, USA, 5–7 October 2016; pp. 415–427.
5. Xu, G.; Xu, C.Z. Prometheus: Online estimation of optimal memory demands for workers in in-memory distributed computation. In Proceedings of the 2017 Symposium on Cloud Computing, Santa Clara, CA, USA, 24–27 September 2017; p. 655.
6. Yan, Y.; Gao, Y.; Chen, Y. Y. Tr-spark: Transient computing for big data analytics. In Proceedings of the Seventh ACM Symposium on Cloud Computing, Santa Clara, CA, USA, 5–7 October 2016; pp. 484–496.
7. Chen, W.; Rao, J.; Zhou, X. Preemptive, low latency datacenter scheduling via lightweight virtualization. In Proceedings of the USENIX Annual Technical Conference (USENIX ATC 17), Santa Clara, CA, USA, 12–14 July 2017; pp. 251–263.
8. Chang-Hong Hsu, Qingyuan Deng and Jason Mars. SmoothOperator: Reducing Power Fragmentation and Improving Power Utilization in Large-scale Datacenters. In Proceedings of the 23rd International Conference on Architectural Support for Programming Languages and Operating Systems(ASPLOS '18), Providence, RI, USA, 13–17 April 2018; pp. 535–548.
9. Gu, L.; Zhou, F.; Meng, H. Study on energy consumption and energy efficiency of data centers in China. *China Energy* **2010**, *32*, 42–45.
10. Tantar, A.A.; Tantar, E. A survey on sustainability in ICT: A computing perspective. In Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation, Vancouver, BC, Canada, 12–16 July 2014; pp. 1213–1220.
11. Zhang, W.; Wen, Y.; Wong, Y.W.; Toh, K.C.; Chen, C.H. Towards joint optimization over ICT and cooling systems in data centre: A survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1596–1616.
12. Xing, B.; Lu, M.; Jin, P.; Huang, G.; Yue, L. Energy efficient query Optimization Technology for Green data Center. *Comput. Res. Dev.* **2019**, *56 (9): 1821-1831.*
13. Song, J.; Sun, Z.; Liu, H.; Bao, Y.; Yu, Y. Research Progress on Energy consumption Optimization of Hybrid Power supply data Center. *J. Comput. Sci.* **2018**, *41*, 36–54.
14. Shen, H. Energy consumption calculation and energy efficiency analysis in data center energy conservation evaluation report. *Eng. Res.* **2019**, 4(15):245-246.
15. Jiang,C.;Wang,Y.;Ou,D.;Qiu,Y.;Li,Y.;Wan,J.;Luo,B.;Shi,W.;Cerin,C.. EASE: Energy Efficiency and Proportionality Aware Virtual Machine Scheduling. In Proceedings of the 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD2018), Paris, France, 24–27 September 2018.
16. Ryckbosch, F.; Polfliet, S.; Eeckhout, L. Trends in server energy proportionality. *Computer* **2011**, *44*, 69–72.
17. Wong, D.; Annavaram, M. Knightshift: Scaling the energy proportionality wall through server-level heterogeneity. In Proceedings of the 2012 45th Annual IEEE/ACM International Symposium on Microarchitecture, Vancouver, BC, Canada, 1–5 December 2012; pp. 119–130.
18. Barroso, L.A.; Hölzle, U. The case for energy-proportional computing. *Computer* **2007**, *40*, 33–37.

19. Buyya, R.; Ranjan, R.; Calheiros, R.N. Modeling and simulation of scalable cloud computing environments and the CloudSim toolkit: Challenges and opportunities. In Proceedings of the 2009 International Conference on High Performance Computing & Simulation, Leipzig, Germany, 21–24 June 2009; pp. 1–11.
20. Calheiros, R.N.; Ranjan, R.; Beloglazov, A.; De Rose, C.A.; Buyya, R. CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Softw. Pract. Exp.* **2011**, *41*, 23–50.
21. Calheiros, R.N.; Ranjan, R.; De Rose, C.A.F.; Buyya, R. CloudSim: A novel framework for modeling and simulation of cloud computing infrastructures and services. *arXiv:0903.2525 [cs.DC]* **2009**. https://doi.org/10.48550/arXiv.0903.2525
22. Garg, S.K.; Buyya, R. Networkcloudsim: Modelling parallel applications in cloud simulations. In Proceedings of the 2011 Fourth IEEE International Conference on Utility and Cloud Computing, Washington, DC, USA, 4–9 July 2011; pp. 105–113.
23. Werner, J.; Geronimo, G.; Westphall, C.; Koch, F.; Freitas, R. Simulator improvements to validate the green cloud computing approach. In Proceedings of the 7th Latin American Network Operations and Management Symposium, Quito, Ecuador, 10–11 October 2011; pp. 1–8.
24. Belalem, G.; Tayeb, F.Z.; Zaoui, W. Approaches to improve the resources management in the simulator CloudSim. In Proceedings of the International Conference on Information Computing and Applications, Tangshan, China, 15–18 October 2010; Springer: Berlin, Germany, 2010; pp. 189–196.
25. Li, X.; Jiang, X.; Ye, K.; Huang, P. DartCSim+: Enhanced cloudsim with the power and network models integrated. In Proceedings of the 2013 IEEE Sixth International Conference on Cloud Computing, Dresden, Germany, 9–12 December 2013; pp. 644–651.
26. Bux, M.; Leser, U. DynamicCloudSim: Simulating heterogeneity in computational clouds. *Futur. Gener. Comput. Syst.* **2015**, *46*, 85–99. https://doi.org/10.1016/j.future.2014.09.007.
27. Guérout, T.; Monteil, T.; Da Costa, G.; Calheiros, R.N.; Buyya, R.; Alexandru, M. Energy-aware simulation with DVFS. *Simul. Model. Pract. Theory* **2013**, *39*, 76–91.
28. Chen, W.; Deelman, E. WorkflowSim: A toolkit for simulating scientific workflows in distributed environments. In Proceedings of the 2012 IEEE 8th International Conference on E-Science, Chicago, IL, USA, 8–12 October 2012.
29. Wickremasinghe, B.. CloudAnalyst: A CloudSim-based tool for modelling and analysis of large scale cloud computing environments. MEDC Project Report, Distrib. Comput. Proj. Csse Dept., University of Melbourne, 2009, Victoria,Australia, pp. 433–659.
30. Wickremasinghe, B.; Calheiros, R.N.; Buyya, R. Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications. In Proceedings of the 2010 24th IEEE International Conference on Advanced Information Networking and Applications, Perth, Australia, 20–23 April 2010; pp. 446–452.
31. Kecskemeti, G. DISSECT-CF: A simulator to foster energy-aware scheduling in infrastructure clouds. *Simul. Model. Pr. Theory* **2015**, *58*, 188–218, https://doi.org/10.1016/j.simpat.2015.05.009.
32. Tian, W.; Zhao, Y.; Xu, M.; Zhong, Y.; Sun, X. A Toolkit for Modeling and Simulation of Real-Time Virtual Machine Allocation in a Cloud Data Center. *IEEE Trans. Autom. Sci. Eng.* **2013**, *12*, 153–161, https://doi.org/10.1109/tase.2013.2266338.
33. Lim, S.H.; Sharma, B.; Nam, G.; Kim, E.K.; Das, C.R. MDCSim: A multi-tier data center simulation, platform. In Proceedings of the 2009 IEEE International Conference on Cluster Computing and Workshops, New Orleans, LA, USA, 31 August–4 September 2009; pp. 1–9.
34. Kliazovich, D.; Bouvry, P.; Khan, S.U. GreenCloud: A packet-level simulator of energy-aware cloud computing data centers. *J. Supercomput.* **2010**, *62*, 1263–1283, https://doi.org/10.1007/s11227-010-0504-1.
35. Hasan, M.S.; Harding, C.; Yu, H.; Griffiths, A. Modeling delay and packet drop in networked control systems using network simulator NS2. *Int. J. Autom. Comput.* **2005**, *2*, 187–194.
36. Kurowski, K.; Oleksiak, A.; Piątek, W.; Piontek, T.; Przybyszewski, A.; Węglarz, J. DCworms—A tool for simulation of energy efficiency in distributed computing infrastructures. *Simul. Model. Pr. Theory* **2013**, *39*, 135–151. https://doi.org/10.1016/j.simpat.2013.08.007.
37. Sriram I. SPECI, a simulation tool exploring cloud-scale data centres. In Proceedings of the IEEE International Conference on Cloud Computing, Bangalore, India, 21–25 September 2009; pp. 381–392.
38. Avetisyan, A.I.; Campbell, R.; Gupta, I.; Heath, M.T.; Ko, S.Y.; Ganger, G.R.; Namgoong, H. Open cirrus: A global cloud computing testbed. *Computer* **2010**, *43*, 35–43.
39. Campbell, R.H.; Gupta, I.; Heath, M.T.; Ko, S.Y.; Kozuch, M.; Kunze, M.; Soh, Y.C. Open cirrus™ cloud computing testbed: Federated data centers for open source systems and services research. *HotCloud* **2009**, *9*, 1–5.
40. Deng, Q.; Meisner, D.; Bhattacharjee, A.; Wenisch, T.F.; Bianchini, R. Coscale: Coordinating CPU and memory system DVFS in server systems. In Proceedings of the 45th Annual IEEE International Symposium on Microarchitecture, Vancouver, BC, Canada, 1–5 December 2012; pp. 143–154.
41. Zhang, Y.; Liu, J.; Kandemir, M. Software-directed data access scheduling for reducing disk energy consumption. In Proceedings of the 2012 IEEE 32nd International Conference on Distributed Computing Systems, Macau, China, 18–21 June 2012; pp. 596–605.