



Evaluating the sustainability of big data centers using the analytic network process and fuzzy TOPSIS

Qingyu Zhang¹ · Shimiao Yang¹

Received: 28 July 2020 / Accepted: 27 October 2020 / Published online: 6 January 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The big data revolution has created data center sustainability problems, whose solutions require the consideration of environmental factors. The purpose of this study is to establish a big data center sustainability evaluation index and provide guidance for sustainable data center construction. This research formulated a big data center sustainability evaluation model that integrates multiple-criteria decision-making methods based on the analytic network process and fuzzy technique for order preference by similarity to an ideal solution (TOPSIS). Furthermore, a case study was used to examine the proposed model. The refrigeration system, layout and ventilation, data center location, data volume, and server power consumption are the five most crucial factors in determining the sustainability level of a big data center. The areas that require further development are the balancing of tasks on different IT equipment, renewable energy use, and waste heat utilization. This research provides a method or guide that can be used by managers when they build new big data centers or upgrade and optimize existing big data centers to make them more sustainable. This study is the first to assess the sustainability of a big data center according to multiple criteria decision-making methods, in which fuzzy theory is applied to evaluate the imprecise and subjective judgments of decision-makers. This study provides a systematic evaluation framework that is based on qualitative and quantitative criteria and comprises the four factors of big data level, equipment level, room level, and data center level. Big data is new oil, but it is not clean oil. It is both a vital driver of economic growth and a source of environmental damage. We need to ensure that big data centers are run in a sustainable way.

Keywords Big data center · Sustainability · Energy consumption · ANP · Fuzzy TOPSIS

Introduction

In recent years, concerns regarding climate change and environmental sustainability have increased. Energy savings and environmental protection are crucial issues that the world faces today (Zhang et al. 2015; Song and Wang 2017; Lucivero 2020). Big data is viewed as an efficient resource that helps reduce material consumption; however, managing and storing such data require considerable energy consumption (McKinsey 2016; Karnama et al. 2019; Rahmani et al. 2020). The big data

revolution has created data center sustainability issues, whose solutions require the simultaneous consideration of economic and environmental factors (Corbett 2018; Herman et al. 2018; Singh and El-Kassar 2019; Kheybari et al. 2020). According to Corbett (2018), the power consumption of IT equipment for storage was 27.8 kWh per TB of data, and the total power consumption of data centers was 46.33 kWh per TB of data per year, which corresponds to approximately 35 kg of CO₂ emissions per TB of data per year. These estimates are disconcerting due to the growing workloads of data center from business needs. The processing of large quantities of data requires massive and efficient computing resources, which raises concerns regarding the harmful environmental effects of big data (Ramli et al. 2017). With the development of the big data industry, high cost, high energy consumption, and high carbon emissions have become increasingly common (Rong et al. 2016; Möbius et al. 2014; Shuja et al. 2016; Lucivero 2020). According to the Climate Group, 76 million metric tons of CO₂ (MtCO₂) were emitted by worldwide data centers in 2002. This figure could reach 259 MtCO₂ by 2020 even after

Responsible Editor: Marcus Schulz

✉ Qingyu Zhang
q.yu.zhang@gmail.com

✉ Shimiao Yang
2170134901@email.szu.edu.cn

¹ College of Management, Research Institute of Business Analytics & Supply Chain Management, Shenzhen University, Shenzhen 518060, China

implementing advanced technologies in virtualization, data center cooling, and power supply (Webb 2008). The data center industry was estimated to account for 1.3% of the world's power consumption and 2% of the USA's power consumption (Nadjahi et al. 2018).

The assessment and management of energy consumption in big data centers are crucial challenges for data center operations. Currently, there is no conceptual framework available to evaluate the sustainability level of big data centers, and most evaluation criteria only include crisp-value ratings, but not qualitative linguistic ratings. To the best of our knowledge, numerous data centers currently use power usage effectiveness (PUE) as the only indicator to assess their energy consumption. The PUE value is closely related to the efficiency of the electricity grid and an accurate calculation of the IT load, and a data center with higher PUE values and higher server utilization is more efficient (Brady et al. 2013). However, IT equipment energy consumption is but a part of the data center's total energy consumption (Whitehead et al. 2014; Ahmad and SMK 2015). To comprehensively evaluate the degree of sustainability of a data center, many factors must be considered in addition to PUE. Data center sustainability depends on not only the room layout, energy sources, and energy consumption of the IT equipment but also the features of the big data processed (e.g., its volume, variety, and velocity). Few studies have attempted to develop a holistic evaluation method for a sustainable data center. Therefore, a flexible and comprehensive evaluation method must be developed, as a new frontier of research, to solve energy problems in data centers and reduce carbon emissions. Therefore, we investigated and reviewed the literature on data center sustainability to develop a four-factor evaluation index system (comprising the factors of big data, equipment level, room level, and data center level) for assessing data center sustainability.

Given the nature of multiple-criteria decision-making, sustainability assessments of large data centers include both qualitative and quantitative criteria. Specifically, experts present linguistic ratings for the qualitative criteria and crisp-value ratings for the quantitative criteria. In many cases, crisp values are not available for the qualitative criteria or the data is insufficiently precise to be used directly in the evaluation process because the judgments and preferences of experts are ambiguous or uncertain, making it impossible to establish an assessment with accurate values (Mehrerjedi 2012). In response, fuzzy theory has been used to address the subjective and imprecise assessments of experts and to capture the ambiguity and vagueness arising from incomplete information in the decision-making process (Zadeh 1965). In this study, by combining qualitative and quantitative measures, a fuzzy technique for order preference by similarity to an ideal solution (TOPSIS) was used to develop a comprehensive performance rating on a data center's sustainability.

The remainder of this paper is structured as follows. Section 2 provides a review of some recent studies on the sustainability and energy consumption evaluation of big data centers. Section 3 introduces the proposed methodology, including the analytic hierarchical process (AHP), the analytic network process (ANP), and its fuzzy extension. In Section 4, we propose a fuzzy multiple-criteria approach for evaluating data center sustainability. In Section 5, a numerical example is used to examine the proposed framework. The conclusions are provided in Section 6.

Literature review

Theoretical background

The increased energy demand from data centers has encouraged operators to take measures to decrease energy consumption and enhance energy efficiency. Accordingly, numerous researchers have attempted to formulate approaches aimed at decreasing energy consumption and enhancing energy efficiency. However, research has indicated that sustainable data centers can be achieved only upon the adoption of a system-wide procedure that involves the holistic management of different data center components. Guitart (2017) proposed a comprehensive management strategy for sustainable data centers that involves introducing energy as a driving force for operation programs. Recent research on the design and architecture of green data centers has mainly focused on the improvement of data center building design, computer room design, floor layout, the cooling and refrigeration system, the electrical system, and the IT equipment. This paper summarized the results of the aforementioned research, noting the four crucial factors of data center level (site selection and energy management), room level (computer room design and heat dissipation), equipment level (IT equipment and electrical systems), and data level (big data resource characteristics and service).

In its site selection and energy management, the data center should be situated where power transmission loss can be reduced, power supply is convenient (Sheme et al. 2018; Daim et al. 2013), and natural cooling can be achieved at low location temperatures. As a new frontier of research, the replacement of traditional energy with renewable energy in data centers has been receiving increasing attention. Data centers using renewable energy have low greenhouse gas emissions due to the decreased use of fossil fuels (Shuja et al. 2016). Oró et al. (2015) proposed numerical models by analyzing renewable energy use in data centers. Furthermore, because most of the electrical supply of data center servers is dissipated as thermal energy, the management of the energy conversion process has become a new and crucial aspect for achieving data center sustainability (Jones and Fleischer 2014). Ideally, waste heat

can be used to offset heating costs (Haywood et al. 2012), and heating facilities in cold regions can partly meet energy demands through the intermittent generation of renewable energy (Woodruff et al. 2014; Lin 2018; Kwon 2020). The main challenge involved in consolidating data center resources for achieving energy efficiency is that maximizing the quality of services conflicts with minimizing the energy consumption of the data center.

With regard to computer room design and heat dissipation, the indoor temperature has a considerable effect on big data center energy consumption. When run at full load, the refrigeration system increases energy consumption while maintaining the temperature in the computer room (Rong et al. 2016). Through suitable cooling methods, the energy efficiency and thermal management of big data centers can be enhanced to various degrees; thus, most data centers have begun to introduce cooling technologies to enhance cooling efficiency (Li et al. 2016; Ni and Bai 2017; Arora and Bala 2020). These technologies are classified into air-cooled, liquid-cooled, and two-phase technologies (Jones and Fleischer 2014). A higher airflow from the layout is crucial to making room temperature more uniform, in addition to enhancing sustainable development and the efficiency of task implementation (Zhang et al. 2014). Technologies for controlling the environment of computer rooms and to ensure that big data centers operate stably and securely are being developed. The aforementioned research developments can contribute to the effective operation and energy savings of data centers.

IT equipment and power distribution systems consume approximately 40% and 15% of the total energy of a data center, respectively (Boru et al. 2015). Industry specifications and the scholarly literature have proposed several metrics for estimating the energy consumption and performance of computing systems. The most important metric focuses on the efficiency of IT facilities, communication systems, and refrigeration and power distribution systems (Fiandrino et al. 2017; Uddin and Rahman 2012; Arora and Bala 2020). PUE is currently the most popular evaluation indicator in many data centers (Capozzoli and Primiceri 2015), measuring the percentage of overall power demand from IT equipment. An increasing awareness of the energy crisis and the importance of IT productivity have shifted managers' view toward a sustainable energy strategy. However, energy-saving IT equipment has had limited use in practice, and many companies lack awareness on whether they should establish large-scale sustainable data centers. Considerably reducing the energy consumption of servers is an important requirement for sustainability (Möbius et al. 2014) because high energy consumption does not entail high performance (Dargie et al. 2011). Data center networks have been evolving to meet the high demand for services and applications (Wang et al. 2013), and the construction of a data center network directly influences the scalability, agility, and power consumption of a data center. Hammadi

and Mhamdi (2014) classified most data center network structures into switch-centric and server-centric topologies. However, no unified and standardized set of ordinances have been formulated for data center design and improvement. Virtual machine migration has also been used to solve load balancing in data centers to eliminate overloaded equipment (Yu et al. 2018). Furthermore, in most recent studies, data centers have widely adopted the multi-objective coevolutionary algorithm and a simulation method for achieving an enhanced green-oriented scheduling strategy (Ham et al. 2015; Lei et al. 2016; Rahmani et al. 2020). Equipment, optimization, integrated design, and other issues related to the data center architecture should be viewed as necessary components of a sustainability evaluation of data centers.

With respect to big data characteristics and information service, massive amounts of data can be regarded as the "raw material" of data centers. Massive data volumes are created daily at record-high rates from heterogeneous provenances (e.g., from healthcare, government, social networks, and commerce) (Oussous et al. 2018; Addo-Tenkorang and Helo 2016; Lucivero 2020), and data centers are the backbone for communication, cloud services, data computing, and other computational services (Weihl et al. 2011). Such data allow for data-driven decision-making and the obtaining of actionable insights. In general, big data refers to a dataset that cannot be captured, stored, conducted, and processed efficiently by normal computers within a reasonable duration (Hashem et al. 2015). Laney (2001) noted volume, velocity, and variety to be the three primary characteristics of big data. In addition to these traits, veracity and value were also proposed as important characteristics of a data center (Kacfeh Emani et al. 2015). These features of big data affect the carbon emissions of a data center. At present, researchers and practitioners have committed to extracting actionable intelligence and discernment from big data in applications such as E-health (Dimitrov 2016), Internet of Things (Zhong et al. 2016), supply chain management and logistics (Sanders and Ganeshan 2018), and retail (Fisher and Raman 2018). A data center provides a platform for data analytics (e.g., descriptive, predictive, exploratory, and prescriptive analytics), allowing for advanced analytic techniques.

Currently, physical-level improvement is the main subject of research in big data (Todorovic and Kim 2014). Many of the practices and measures adopted by firms to successfully improve the energy consumption of data centers involve improvements to IT equipment and the cooling system (Priyadumkol and Kittichaikarn 2014; Kim et al. 2017; Kunkel et al. 2019). However, few studies have addressed the sustainable development of big data itself. Big data may not convey valuable information, and additional data do not entail better decisions (Corbett 2018). Approximately 90% of the data generated on the market is never used, and 60% of these data become useless within milliseconds (Johnson

2015). In this study, we incorporated the features of data in evaluating the sustainability of a big data center. Moreover, although many studies have examined the energy consumption of traditional data centers (e.g., PUE), none has systematically appraised the sustainability of big data centers. Hence, this study formulated a comprehensive method that evaluates the source, consumption, and output (e.g., waste heat utilization) of energy in a data center through accounting for the four aforementioned factors. Furthermore, in this study on the sustainability of a big data center, fuzzy ANP was used to analyze responses to the qualitative and quantitative criteria adopted. Generally, big data centers should be constructed and developed in a sustainable and cautious manner.

Sustainability evaluation

Big data have been characterized as 5 Vs: volume, variety, velocity, veracity, and value (Cheng et al. 2017). The details of five aspects are as follow: (1) volume: the amount of data is huge and the size is enormous; (2) variety: the range of data types and sources is large (e.g., multiple sources with multiple dimensions) and data is mixed, heterogeneous, and unstructured; (3) velocity: the speed of data changes is high and data is accumulated rapidly; (4) veracity: the data quality is unreliable due to data inconsistency, incompleteness, and data accuracy is uncertain; (5) value: the value refers to the insights and benefits from data, and maximizing value is to gain insights in real time (e.g., fraud detection). Big data centers aim to get the insights or benefits from the massive amount of changing data. The huge amount and rapidly varying data increases the hardness of getting benefits from them. Even for a simple search operation, the problem with big data is much more difficult than a problem with ordinary data. Big data cannot be handled by typical database, normal computers, or traditional software. It needs huge parallel processing power of computer clusters. Big data look like chaotic, but some hidden knowledge can be discovered. However, the operations of big data centers consume huge energy and generate enormous carbon emissions. Out of the 5 Vs, the first three (i.e., volume, variety, and velocity) are more related to the sustainability of data centers.

Researchers have studied the energy consumption and performance of data centers from different aspects, such as computing power, cooling, and network-related factors. Many companies are still using PUE-based assessment methods. However, sustainability evaluation involves the examination of various characteristics of a data center; among these characteristics, uncertain criteria are challenging to measure. Hence, few studies have successfully established a holistic system for evaluating data center sustainability. Through an extensive literature review and discussions with experts in the data center industry, the criteria presented in Table 1 were determined to be important to data center sustainability. The criteria were divided into four categories: big data level, equipment level, room level, and data center level. The big

data level comprises criteria on the volume, velocity, and variety of big data. The equipment level comprises criteria on PUE, the network, storage devices, power sourcing equipment, the server, and the quantity of IT racks. The room level comprises criteria on room monitoring and management, layout and ventilation, the computer room environment, and the refrigeration system. The data center level comprises criteria on renewable energy, waste heat utilization, data center location, and energy consumption per unit area.

Methods

An integrated approach was proposed as effective to evaluate the sustainability of big data centers by combining the ANP and fuzzy TOPSIS. ANP is used to weight the relative importance of the dimensions and subdimensions; then, performance scores and weights are combined using fuzzy TOPSIS. This approach is specifically useful for dealing with the situations where the evaluations are uncertain and imprecise. This integration is applied successfully to the complex big-data-center-sustainability problem of the vague and imprecise nature of linguistic assessments, where the experts' comparisons are denoted as fuzzy numbers.

ANP

The AHP was first proposed by Saaty (1980) for hierarchically structured decision problems. Based on expert judgments, this method is implemented by comparing measurements obtained through absolute scales on tangible and intangible criteria (Saaty 1980). However, many real decision problems are not hierarchically structured: their structure includes interdependent relationships and feedback among the components and decision levels. Therefore, the ANP is more appropriate to such problems, where a network is treated as having clusters of elements rather than a hierarchy (Saaty and Vargas 2006). The ANP has been widely applied in many decision-making problems in the last decade (Seyhan and Mehpare 2010).

Fundamental scale

To balance between competing multiple objectives and criteria, decisions are usually made qualitatively to determine the numerical value that should represent relative importance. As in the AHP, we should make pair-wise comparisons using a scientific and carefully designed approach. The fundamental scale used for decision-making is presented in Table 2. In this research, the ANP was applied to weight the relationships among the main criteria and subcriteria with respect to the goal.

Table 1 List of criteria for evaluating sustainability

Criteria	Subcriteria	Measures	Sources
Big data	Volume	(1) Capacity of the data center server for hosting, cloud computing, and providing Internet access. (2) Utilization = $\text{Capacity}_{\text{Used}} / \text{Capacity}_{\text{Total}}$.	Laney 2001; Weihl et al. 2011; Oussous et al. 2018; Chen et al. 2014; Kacifah Emami et al. 2015; Hashem et al. 2015; Arora and Bala 2020
	Velocity	Velocity refers to the timeliness and continuity of big data. Tenant data should be conducted rapidly to better extract useful information and maximize commercial value.	
	Variety	Data collected through various types of sensors for serving different types of clients.	
	Power sourcing equipment	Power distribution cabinet and UPS. A UPS is an alternate or backup source of power that is used to ensure continued operation in the event of a power outage.	
Equipment level	PUE	Ratio between the total power consumed by a data center and the total power delivered to its IT equipment. PUE is necessarily greater than 1, and higher values indicate higher energy efficiency.	Dargie et al. 2011; Uddin and Rahman 2012; Wang et al. 2013; Hammadi and Mhamdi 2014; Möbius et al. 2014; Boru et al. 2015; Capozzoli and Primiceri 2015; Lei et al. 2016; Fiandrino et al. 2017; Yu et al. 2018; Lucivero 2020
	Network	The network can be assigned to any service and server for dynamic resource allocation.	
Room level	Storage devices	Reduce time disk access and capitalize on energy-efficient storage hardware (power, capacity, performance, and dependability).	Jones and Fleischer 2014; Zhang et al. 2014; Rong et al. 2016; Fiandrino et al. 2017; Ni and Bai 2017; Li et al. 2016; Rahmani et al. 2020
	Server	Power consumption and efficiency of the server depends on CPU frequency, packet size, and transmission rate.	
	Quantity of IT racks	Avoid overpurchasing unnecessary or inefficient equipment when using servers, storage, and networking.	
	Refrigeration system	The refrigeration system's efficiency should be maximized by ensuring strict insulation; the system should be waterproof, leakproof, and dustproof.	
	Layout and ventilation	The raised floor plenum, alternating hot and cold aisle, and computer room air-conditioning airflow.	
	Room monitoring and management	Firefighting system, device control, remote monitoring, monitoring management network, and illumination.	
	Computer room environment	Air-conditioning group control, conditioning humidification, electrostatic, radiation, and clean air.	
	Energy consumption per unit area	Quotient of the energy consumption of a data center and the total data center area.	
	Data center location	Sufficient power supply capacity, low price of electricity supply, and low annual average temperature.	
	Waste heat utilization	Resulting data center collects and reuses waste heat.	
Renewable energy	Natural climatic conditions and renewable energy can be obtained depending on the geographical location and percentage of renewable energy used to power data centers.	Haywood et al. 2012; Daim et al. 2013; Woodruff et al. 2014; Oró et al. 2015; Shuja et al. 2016; Guitart 2017; Sheme et al. 2018; Kwon 2020	

Table 2 Fundamental evaluation scale

Numerical importance	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance
2, 4, 6, and 8	Intermediate values
Reciprocals of the aforementioned nonzero numbers	If activity i has one of the aforementioned numbers assigned to it relative to activity j , then j has the reciprocal of that number relative to i .

Main steps of the ANP

Model construction The purpose of the data center evaluation problem should be determined, and the elements affecting the goal of the model must be divided into main criteria and subcriteria. This division results in a three-layer hierarchical model structure. According to the opinions of decision-makers, the relationships among the elements in each cluster are marked by arrows. Finally, the alternative data centers are incorporated into the model as the fourth layer (e.g., Fig. 2).

Pair-wise comparison The elements of every cluster are compared singly according to their importance for their control criteria. Judges then evaluate the pair-wise comparisons. For every pair-wise comparison matrix, the consistency ratio (CR) is used to evaluate for logical consistency. A CR value < 0.1 indicates a consistent pair-wise comparison matrix.

$$CR = \frac{\lambda_{max} - n}{(n - 1) * RI}$$

Where λ_{max} = the average of normalized weighted sum vector, n = number of criteria; RI = consistency index of a randomly generated pairwise comparison matrix.

Formation of the supermatrix To obtain the global priorities, the local preference is calculated in the appropriate columns of the supermatrix. This forms a partitioned matrix. Every section represents the contact between clusters. However, the supermatrix is unweighted and must be transferred into relevant limiting priorities. For this purpose, a weighted supermatrix must be obtained by normalizing each column of the unweighted supermatrix.

Acquiring the weights of criteria The weighted supermatrix is converted to a limiting supermatrix, where the relative weights of the elements can then be obtained from the rows of the limiting supermatrix. Eventually, the overall score S_j of a decision alternative is shown as follows.

$$S_j = \sum_i w_i r_{ij}$$

Where w_i = the weight for criterion i , r_{ij} = the rating for criterion i and decision alternative j .

Fuzzy set theory

Decision-making is difficult in an ambiguous and uncertain environment. Zadeh’s fuzzy set theory (Zadeh 1965) can be used to better regulate ambiguity and uncertainty. Fuzzy set theory can improve the synthesis and rationalization of the decision-making process (Chen 2000), and it has been applied to resolve ambiguities in human perception and decision-making. The use of fuzzy sets has enabled the importance of criteria to be evaluated in linguistic terms.

Establishing the fuzzy number

To enable decision-makers to precisely assess criteria, a triangular fuzzy number is commonly introduced in most fuzzy applications (Lan 2016). Figure 1 illustrates a triangular fuzzy number M (Shaw et al. 2012). We represent a fuzzy number as (a, m, b) , and the function of membership is presented in Eq. (1). The lower and upper bounds of the fuzzy number M are a and b , respectively, where m is the mode of M (Lee et al. 2009).

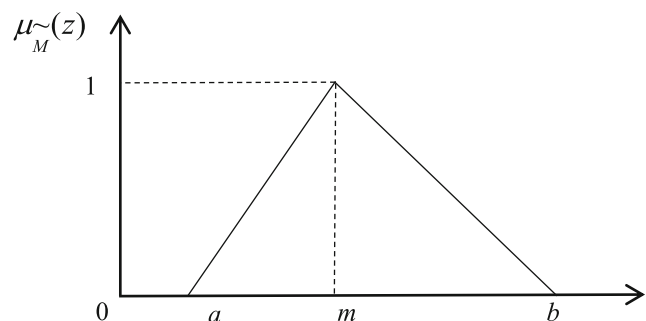


Fig. 1 Triangular fuzzy number

$$\mu_{\tilde{M}}(z) = \begin{cases} (z-a)/(m-a), & a \leq z \leq m \\ (b-z)/(b-m), & m \leq z \leq b \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Let two triangular fuzzy numbers M_1 and M_2 be parameterized by (a_1, m_1, b_1) and (a_2, m_2, b_2) , respectively. We propose the following order of operations for triangular fuzzy numbers.

$$\begin{aligned} M_1 + M_2 &= (a_1, m_1, b_1) + (a_2, m_2, b_2) \\ &= (a_1 + a_2, m_1 + m_2, b_1 + b_2) \end{aligned} \quad (2)$$

$$\begin{aligned} M_1 \times M_2 &= (a_1, m_1, b_1) \times (a_2, m_2, b_2) \\ &= (a_1 a_2, m_1 m_2, b_1 b_2) \end{aligned} \quad (3)$$

$$\begin{aligned} M_1 \div M_2 &= (a_1, m_1, b_1) \div (a_2, m_2, b_2) \\ &= (a_1/a_2, m_1/m_2, b_1/b_2) \end{aligned} \quad (4)$$

The distance between M_1 and M_2 is defined as follows (Krohling and Campanharo 2011).

$$d(M_1, M_2) = \sqrt{[(a_1 - a_2)^2 + (m_1 - m_2)^2 + (b_1 - b_2)^2]} / 3. \quad (5)$$

Determining the linguistic variables

Linguistic terms are part of the subjective labels of the linguistic variable. We defined the value of a linguistic variable to be a word or phrase. In this study, nine basic linguistic terms were used in the pair-wise comparisons of the sustainability of two data centers (Gumus 2009). These terms are listed in Table 3.

Fuzzy TOPSIS

TOPSIS is generally used to solve ranking problems with a finite set of alternatives. The method depends

Table 3 Linguistic variables for pair-wise comparisons of criteria

Fuzzy number	Linguistic	Scale of fuzzy number
9	Perfect (Pe)	(8, 9, 10)
8	Absolute (A)	(7, 8, 9)
7	Very good (VG)	(6, 7, 8)
6	Fairly good (FG)	(5, 6, 7)
5	Good (G)	(4, 5, 6)
4	Preferable (Pr)	(3, 4, 5)
3	Not bad (NB)	(2, 3, 4)
2	Weak advantage (WA)	(1, 2, 3)
1	Equal (E)	(1, 1, 1)

on the principle that the selected alternative should have the shortest distance from the positive-ideal solution and the farthest distance from the negative-ideal solution (Sun 2010; Gumus 2009). In this study, we evaluated the sustainability of data centers using TOPSIS. The steps in the fuzzy TOPSIS procedure are as follows (Kuo et al. 2007; Önüt and Soner 2008; Kaya and Kahraman 2011).

Step 1: Weight the evaluation criteria. In this study, fuzzy ANP was used to detect the fuzzy preference weights.

Step 2: Construct the fuzzy decision matrix and select suitable linguistic variables for the alternatives in relation to the criteria. Specifically,

$$\tilde{D} = \begin{matrix} & C_1 & C_2 & \cdots & C_q \\ \begin{matrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{matrix} & \begin{bmatrix} \tilde{z}_{11} & \tilde{z}_{12} & \cdots & \tilde{z}_{1q} \\ \tilde{z}_{21} & \tilde{z}_{22} & \cdots & \tilde{z}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{z}_{p1} & \tilde{z}_{p2} & \cdots & \tilde{z}_{pq} \end{bmatrix} \end{matrix}, \quad (6)$$

where \tilde{z}_{ij}^k is the weight given by the k th expert and describes the performance of alternative A_i with respect to criterion C_j . The parameter \tilde{z}_{ij}^k is such that $\tilde{z}_{ij}^k = (a_{ij}^k, m_{ij}^k, b_{ij}^k)$.

Step 3: Construct the normalized fuzzy decision matrix. The matrix is expressed as follows:

$$R = [\tilde{r}_{ij}]_{p \times q}, \quad i = 1, 2, \dots, p; j = 1, 2, \dots, q. \quad (7)$$

The normalization method can be implemented using the following equations:

$$\begin{aligned} r_{ij} &= (a_{ij}/b_j^+, m_{ij}/b_j^+, b_{ij}/b_j^+) \text{ and } b_j^+ = \max_i b_{ij} \quad (8) \\ &\text{(when } r_{ij} \text{ is a benefit criterion)} \end{aligned}$$

$$\begin{aligned} r_{ij} &= (a^-/b_{ij}, a^-/m_{ij}, a^-/a_{ij}) \text{ and } a_j^- = \min_i a_{ij} \quad (9) \\ &\text{(when } r_{ij} \text{ is a cost criterion)} \end{aligned}$$

Step 4: Calculate the weighted normalized fuzzy decision matrix U by multiplying the fuzzy decision matrix (R) with the evaluation criteria weights (W_j).

$$U = [u_{ij}]_{p \times q}, \quad i = 1, 2, \dots, p; j = 1, 2, \dots, q$$

The weighted normalized value u_{ij} is calculated using Eq. (10).

$$u_{ij} = r_{ij} \odot w_j \quad (10)$$

Step 5: Define the fuzzy positive-ideal solution (FPIS) and fuzzy negative-ideal solution (FNIS). The component u_{ij} is a normalized positive triangular fuzzy number whose limits are within $[0, 1]$. The FPIS S^+ and FNIS S^- can be determined using the following.

$$S^+ = (u_1^+, \dots, u_j^+, \dots, u_q^+) \quad (11)$$

$$S^- = (u_1^-, \dots, u_j^-, \dots, u_q^-) \quad (12)$$

Step 6: Calculate the Euclidean distance of each alternative from S^+ to S^- .

$$D_i^+ = \sum_{j=1}^q d(u_{ij}, u_j^+), \quad i = 1, 2, \dots, p; j = 1, 2, \dots, q \quad (13)$$

$$D_i^- = \sum_{j=1}^q d(u_{ij}, u_j^-), \quad i = 1, 2, \dots, p; j = 1, 2, \dots, q \quad (14)$$

d_u represents the distance between two fuzzy numbers, which can be calculated using Eq. (5).

Step 7: Calculate the relative closeness of each alternative to the ideal solution. The relative closeness is as follows.

$$CC_i = D_i^- / (D_i^+ + D_i^-) \quad (15)$$

Step 8: Rank the alternatives in descending order according to their relative closeness. The basic principle is that the best alternative has the highest closeness coefficient and is thus closest to the FPIS and farthest from the FNIS.

Proposed model

The proposed model comprises four basic stages. In the first stage, the objective of evaluation is determined. This determination is crucial for applied big data research. In the second stage, the criteria to be used in the model are identified through a literature review and through consultations with experts. The problem examined in this study had a decision-making structure that was hierarchical and networked, as illustrated in Fig. 2. For three data centers, the network of their performance assessment can be clearly visualized. We selected four main criteria (B1 to B4) and 17 subcriteria (C1 to C17)

for constructing the model. Among the 17 subcriteria, C4 and C14 are cost criteria, whereas the other subcriteria are benefit criteria.

After constructing the framework, we calculated the individual priority weights for each criterion using the ANP. Subsequently, the relative weights of the criteria in the supermatrix were obtained and the matrix of the weighted decision was constructed. In the last stage, alternatives were evaluated and ranked using fuzzy TOPSIS. By measuring, in linguistic terms, the decision-makers' opinions on the performance of alternatives, fuzzy TOPSIS was used to rank the data centers.

Application of the proposed model

To evaluate alternative solutions and help managers of big data centers improve sustainability, data center sustainability was evaluated using the proposed model in a case study. The alternatives (A1, A2, and A3) are three big data centers located in universities, one in north China (Beijing), one in south China (Shenzhen), and one in central China (Wuhan). The main service items of the big data centers include scientific research, school–enterprise cooperation, managed rental services, bandwidth access, and security services. An increase in data processing volume and energy consumption has resulted in considerable pressure on the implementation of green practices in traditional big data centers. The case study was used to illustrate how data center sustainability is affected when different aspects are integrated into the evaluation problem.

Calculation of the criteria weights

After developing the research evaluation framework, the Super Decisions software was used to calculate the criteria weights according to the ANP. Four groups of pair-wise matrices were determined for the interdependent relationships and relative importance among the main criteria, subcriteria, and alternatives. Expert opinions for the pair-wise comparison were solicited from professors and practitioners in the fields of big data and data analytics. The first pair-wise matrices were used for assessing the influence of the main criteria on the evaluation objectives, and the comparison results of the criteria are presented in Table 4. For the second group of matrices, we analyzed the interdependence between B1 (Big data), B2 (Equipment), B3 (The room), B4 (Data center), and their subcriteria. The last group of pair-wise matrices describes the effect of the subcriteria (C4 to C17) on the alternatives (A1 to A3). The weighted supermatrix is presented in Table 5, and the limited supermatrix is presented in Table 6. From the results of the limited supermatrix, we obtained the relative weights of each subriterion.

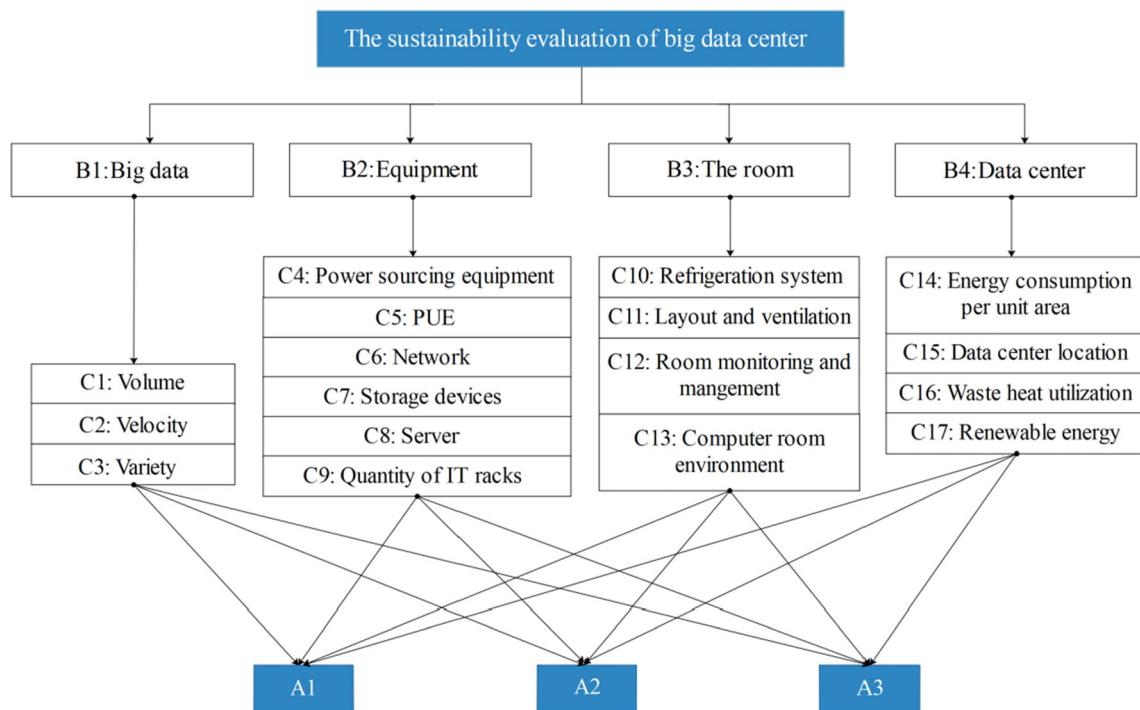


Fig. 2 Multiple-criteria decision-making model for evaluating the sustainability of big data centers

Evaluation of alternatives and determination of the final rank

This case study focused on evaluating the sustainability of three data centers. Three experts used a nine-item scale (Table 4) to judge the relative importance of the selection criteria when performing the pair-wise comparisons. The opinions of all experts were of equal weight, and the mean value was used to represent the overall fuzzy value of the experts’ collective judgment for the same evaluation dimensions. Subsequently, linguistic variables, such as “Perfect,” “Absolute,” “Very good,” “Fairly good,” “Good,” “Preferable,” “Not bad,” “Weak advantage,” and “Equal” were used to evaluate the performance of the data center with respect to the criteria in Table 4. The decision matrix was also constructed by comparing the alternatives in relation to each subcriterion (see Tables 5 and 7). The weight of each subcriterion is also presented in Fig. 3. The assessments of data center performance by the three experts are presented in Table 7.

First, the decision matrix should be normalized with the triangular fuzzy numbers using Eqs. (8) and (9), which include benefit and cost criteria. For criteria C1 of the alternative A1, the normalized value for r_{11} is as follows:

$$z_{11} = (4 + 6 + 7, 5 + 7 + 8, 6 + 8 + 9)/3$$

$$= (5.67, 6.67, 7.67) \tag{16}$$

$$r_{11} = z_{11}/9 = (0.63, 0.74, 0.85) \tag{17}$$

After we obtained the fuzzy evaluation matrix, the fuzzy weighted decision table was calculated. We then obtained the weighted evaluation matrix, as presented in Eq. (10), using the criteria weights computed with the ANP (Table 5). The fuzzy weighted decision matrix is presented in Table 8.

$$u_{11} = r_{11} \times w_{11} = (0.63, 0.74, 0.85) \times 0.066$$

$$= (0.042, 0.049, 0.056) \tag{18}$$

Table 4 Comparison results for the criteria and the criteria’s relative weights

	B1:Bigdata	B2:Equipment	B3:The room	B4:Data center	Weight
B1:Big data	1	2	3	6	0.4564
B2:Equipment	1/2	1	3	7	0.3334
B3:The room	1/3	1/3	1	5	0.1609
B4:Data center	1/6	1/7	1/5	1	0.0493
					CR = 0.0622

Table 5 Weighted supermatrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
C1	0.000	0.100	0.167	0.099	0.103	0.067	0.144	0.067	0.094	0.112	0.000	0.000	0.106	0.066	0.000	0.000	0.000
C2	0.167	0.000	0.167	0.039	0.065	0.029	0.020	0.067	0.037	0.064	0.000	0.000	0.028	0.083	0.000	0.000	0.000
C3	0.167	0.100	0.000	0.062	0.082	0.154	0.086	0.067	0.119	0.024	0.000	0.000	0.067	0.052	0.000	0.000	0.000
C4	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.020	0.000	0.000	0.000
C5	0.000	0.007	0.000	0.200	0.000	0.050	0.029	0.013	0.012	0.009	0.012	0.018	0.010	0.015	0.250	0.000	0.017
C6	0.028	0.015	0.028	0.000	0.029	0.000	0.154	0.065	0.035	0.028	0.024	0.058	0.033	0.025	0.000	0.000	0.045
C7	0.153	0.092	0.215	0.000	0.019	0.000	0.000	0.086	0.053	0.021	0.040	0.043	0.020	0.017	0.000	0.000	0.023
C8	0.073	0.038	0.090	0.000	0.073	0.000	0.000	0.000	0.150	0.085	0.106	0.127	0.064	0.066	0.000	0.000	0.093
C9	0.064	0.049	0.000	0.000	0.129	0.200	0.067	0.037	0.000	0.043	0.068	0.087	0.073	0.057	0.000	0.000	0.071
C10	0.000	0.128	0.000	0.123	0.000	0.000	0.000	0.141	0.000	0.000	0.188	0.039	0.019	0.062	0.250	0.000	0.188
C11	0.000	0.042	0.000	0.045	0.000	0.250	0.250	0.042	0.250	0.150	0.000	0.205	0.125	0.039	0.000	0.000	0.000
C12	0.000	0.020	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.056	0.000	0.000	0.000	0.000
C13	0.000	0.010	0.000	0.018	0.000	0.000	0.000	0.017	0.000	0.050	0.063	0.089	0.000	0.099	0.000	0.000	0.063
C14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042
C15	0.000	0.200	0.000	0.040	0.036	0.000	0.000	0.050	0.000	0.133	0.100	0.000	0.000	0.067	0.000	0.500	0.208
C16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000
C17	0.000	0.000	0.000	0.160	0.214	0.000	0.000	0.150	0.000	0.067	0.082	0.000	0.000	0.130	0.189	0.000	0.000

According to Table 8, we normalized the elements U_{ij} , and $\forall i, j$ as positive triangular fuzzy numbers, with values within $[0, 1]$. Therefore, the FPIS and FNIS were assigned as $u_i^+ = (1, 1, 1)$ and $u_i^- = (0, 0, 0)$ for the benefit criterion, respectively, and $u_i^+ = (0, 0, 0)$ and $u_i^- =$

$(1, 1, 1)$ for the cost criterion, respectively. The distance between the FPIS and FNIS of each criterion of alternatives was calculated using Eq. (5).

$$d_{u_{11}}^+ = \sqrt{[(1-0.041)^2 + (1-0.049)^2 + (1-0.056)^2]}/3 = 0.951 \quad (19)$$

Table 6 Limited supermatrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
C1	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066
C2	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
C3	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058
C4	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
C5	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
C6	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
C7	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057
C8	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060
C9	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
C10	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079
C11	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075
C12	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
C13	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
C14	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
C15	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072
C16	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018
C17	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057
Weight	0.066	0.050	0.058	0.003	0.036	0.037	0.057	0.060	0.052	0.079	0.075	0.003	0.019	0.003	0.072	0.018	0.057

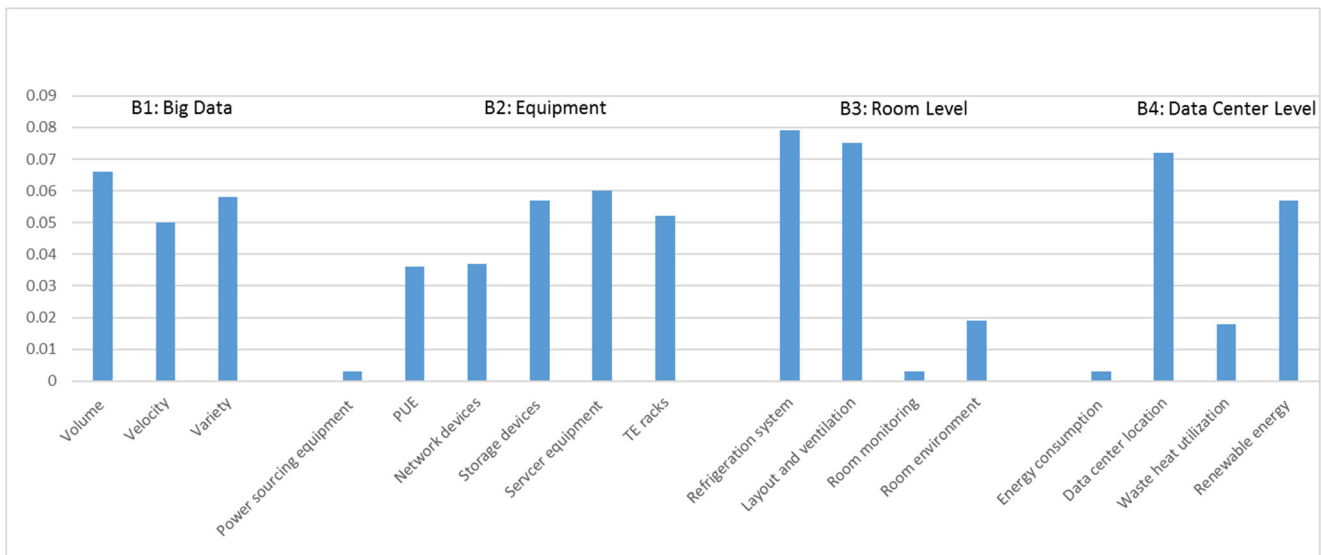


Fig. 3 Weights of subcriteria

$$d_{u_{11}}^- = \sqrt{[(0-0.041)^2 + (0-0.049)^2 + (0-0.056)^2]}/3 = 0.049 \quad (20)$$

The relative closeness of alternatives between the FPIS and FNIS for each criterion of alternatives could be determined by adding the results obtained from Eqs. (13) and (14). Similarly, computations were performed for the other alternatives, and the results of fuzzy TOPSIS analyses are presented in Table 9. The last stage involved solving for the similarities to the ideal

solution using Eq. (15). As indicated by the data in Table 9, the final ranking was A1 > A3 > A2.

Results and discussion

The sustainability of big data centers was evaluated and CC_i values were obtained. According to the CC_i values, the alternatives, in descending order, were A1, A3, and A2. A1 was the best performing data center, having the highest CC_i of 0.1402. The weight of each element was obtained from a

Table 7 Ratings of the three alternatives under 14 criteria

		A 1			A 2			A 3		
		E1	E2	E3	E1	E2	E3	E1	E2	E3
B1	C1	G	VG	A	NB	G	EG	Pr	Pr	VG
	C2	FG	G	Pr	FG	Pr	NB	G	NB	WA
	C3	Pr	VG	A	VG	A	FG	VG	G	G
B2	C4	A	FG	G	NB	NB	E	G	FG	NB
	C5	Pr	G	Pr	FG	A	G	NB	NB	WA
	C6	G	Pr	FG	Pr	WA	NB	Pr	G	NB
	C7	FG	A	A	WA	Pr	G	G	Pr	Pr
	C8	NB	WA	NB	NB	Pr	NB	VG	FG	VG
	C9	FG	VG	A	Pr	Pr	G	WA	NB	Pr
B3	C10	G	FG	FG	FG	A	VG	A	VG	VG
	C11	NB	NB	Pr	FG	VG	G	G	Pr	NB
	C12	A	Pe	VG	Pr	FG	FG	WA	NB	Pr
	C13	FG	VG	A	VG	VG	A	G	NB	Pr
B4	C14	A	G	VG	FG	Pr	G	NB	E	NB
	C15	Pr	G	G	WA	WA	E	E	WA	NB
	C16	VG	A	VG	G	G	Pr	Pr	FG	G
	C17	FG	VG	A	WA	NB	WA	Pr	NB	NB

Table 8 Weighted evaluation matrix

	A1	A2	A3
C1	(0.042,0.049,0.056)	(0.034,0.044,0.053)	(0.033,0.042,0.050)
C2	(0.029,0.036,0.043)	(0.024,0.031,0.038)	(0.020,0.028,0.072)
C3	(0.041,0.046,0.052)	(0.039,0.045,0.052)	(0.034,0.041,0.048)
C4	(0.002,0.002,0.002)	(0.002,0.001,0.001)	(0.001,0.001,0.001)
C5	(0.013,0.017,0.021)	(0.021,0.025,0.029)	(0.001,0.011,0.015)
C6	(0.016,0.021,0.025)	(0.008,0.012,0.016)	(0.012,0.016,0.021)
C7	(0.040,0.042,0.053)	(0.017,0.023,0.030)	(0.021,0.027,0.034)
C8	(0.011,0.018,0.024)	(0.016,0.022,0.029)	(0.038,0.044,0.051)
C9	(0.035,0.040,0.046)	(0.019,0.025,0.031)	(0.012,0.017,0.023)
C10	(0.041,0.050,0.059)	(0.053,0.061,0.070)	(0.056,0.064,0.076)
C11	(0.019,0.028,0.036)	(0.042,0.050,0.058)	(0.025,0.033,0.042)
C12	(0.002,0.003,0.003)	(0.001,0.002,0.002)	(0.001,0.001,0.001)
C13	(0.013,0.015,0.017)	(0.013,0.015,0.018)	(0.006,0.008,0.011)
C14	(0.002,0.002,0.003)	(0.001,0.002,0.002)	(0.001,0.001,0.001)
C15	(0.029,0.037,0.045)	(0.008,0.013,0.019)	(0.011,0.016,0.021)
C16	(0.013,0.015,0.017)	(0.007,0.009,0.011)	(0.008,0.010,0.012)
C17	(0.047,0.041,0.036)	(0.043,0.024,0.017)	(0.024,0.017,0.013)

Table 9 Result of the fuzzy TOPSIS analyses

Alternatives	D^+	D^-	CC_i	Rank
A1	14.617	2.383	0.1402	1
A2	14.652	2.343	0.1381	3
A3	14.634	2.366	0.1391	2

previous study, which used the ANP to evaluate the pair-wise comparison matrix (Table 6). In the evaluation of data center sustainability using the ANP, C10 (refrigeration system), C11 (layout and ventilation), C15 (data center location), C1 (data volume), and C8 (server) were determined to be the five most important criteria. Conversely, C16 (waste heat utilization), C14 (renewable energy), C12 (room monitoring and management), and C4 (power sourcing equipment) were determined to be the four least important criteria. Finally, we calculated the CR of the pair-wise comparison matrix to be 0.0622, which was less than 0.1, indicating that consistent weights were obtained and could be used in the evaluation process.

Rong et al. (2016) provided an energy consumption distribution diagram for data centers, in which the two most important factors were refrigeration system energy consumption (40%) and storage device energy consumption (40%). The results of this study's proposed method indicated that the refrigeration system and servers are the main contributors of energy consumption in a data center, consistent with Rong et al. (2016). As expected, the main concerns in a sustainable data center are related to refrigeration and hardware, particularly the server. Moreover, in the construction of data centers, a suitable data center location as well as layout and ventilation are used to lower cooling costs. Another factor that requires attention is the volume of big data, which is a significant indicator of the data handling capabilities of a data center. Data centers must reasonably distribute data processing tasks to appropriate IT equipment to avoid idle hosts, energy wastage, or overload due to large data volumes. Furthermore, sustainable energy techniques have been a challenge to exploit for the industry. Nevertheless, data center operators should pay more attention to renewable energy use and waste heat utilization, which have considerable potential for improving data center sustainability.

Like any new technologies, big data have sweeping externalities, both good and bad. The positive aspects of big data are real-time monitoring or periodic updating of water, electricity, and other resources at greater granularity, resulting in a greater variety, volume, and velocity of data. By sharing such big data and information with managers, Disney has reduced its electricity usage by nearly billion kWh annually. Caterpillar's customer found out that using more generators at lower output is better than using fewer ones at maximum power (Corbett 2018).

Big data might help people make better decisions; however, large-scale big data centers have their concomitant downsides and uninvited or detrimental consequences—they can consume too much energy or generate too many carbon emissions. From the perspective of carbon emissions, data volume is more important than variety and velocity based on our study. But the large volume of data might not be the right data for decision making. Sometimes, we need to ask what data is not being collected (Sachs 2012). Even there are big volumes of data available for analysis, they might not be correct—veracity concerns. How these data are collected or reported needs to be verified. The more accurate the carbon emission data, the lower emissions a big data center can achieve (Melville et al. 2017). In addition, when big volume data become available, we might be past the prevention time and into the risk mitigation time, like the case of COVID-19 pandemic. Thus, we must also concentrate on making decisions with multi-criteria with little rather than a large volume of data.

What is alarming is that most stored data are considered as waste or dark data. These gigantic amounts of data just safely stay there, but only 0.5% of the data in the world is used and 99.5% is dark data, of which 32% is trivial, obsolete, and redundant (Cohen 2018). A substantial amount of energy use for big data can be avoided if the dark data can be reduced. Thus, we must find a way to curtail the unrestrained data growth including being collected, reported, stored, distributed and disposed of.

Big data can help reduce material or energy consumption, but running big data centers entails physical processes (e.g., server, equipment, and room) consuming energy. The biggest data centers in the world are ever more depending on renewable energy. Apple's and Google's data centers run on 100% renewable energy (Corbett 2018; Kwon 2020). Firms that invest in green IT achieve lower energy consumption and higher profits (Khuntia et al. 2018). Thus, it is necessary to start gauging and abating the environmental costs of big data centers by investing in renewable energy.

Conclusions

This study proposes a data center sustainability evaluation model in which multiple-criteria decision-making methods are integrated using the ANP and fuzzy TOPSIS. The ANP is used to obtain the relative criteria weights for determining their interdependent relationships, and fuzzy TOPSIS is used to rank the performance of alternatives. The applicability of the proposed model was examined using a case study.

Numerous studies have examined technologies for reducing the energy consumption of data centers. However, few theoretical studies have investigated how energy consumption can be reduced by managing data characteristics, energy input (power source), and energy output (location and waste heat

treatment). This study proposes a comprehensive strategy, based on a holistic architecture, for the evaluation of the energy consumption of a data center. The results suggest that designers of a data center should pay attention to the center's carbon footprint and the various factors—such as data characteristics, IT equipment, refrigeration system, and waste heat utilization—affecting energy consumption. Refrigeration system, layout and ventilation, data center location, data volume, and server power consumption were determined to be the five most significant elements in evaluating data center sustainability. The areas that require further development are renewable energy use, waste heat utilization, and how tasks should be optimally allocated to various IT equipment.

To the best of our knowledge, this study is the first to assess data center sustainability through multiple-criteria decision-making methods, where fuzzy theory is used to evaluate the imprecise and subjective judgments of decision makers. This study formulated a systematic evaluation framework based on qualitative and quantitative criteria and comprising the four factors of big data level, equipment level, room level, and data center level. This framework provides a method that managers can use to enhance sustainability when constructing new data centers or upgrading and optimizing existing ones. The managers of data centers must strive to optimize the refrigeration system, layout and ventilation system, data center location, and arrangement of tasks to increase the energy efficiency of IT equipment. They should also better understand clean energy use and waste heat utilization. Moreover, the research results indicate that achieving sustainability and energy consumption reduction for a data center is a systematic problem influenced by factors such as the efficiency of equipment, airway layout, location, data characteristics, and renewable energy. Big data is like unclean oil, which is not only a powerful force for economic development, but also a root cause of environmental harm. We need to make certain that big data centers operate in a sustainable way.

Acknowledgements The authors thank the Editor and anonymous reviewers for their valuable contributions that improved this manuscript.

Authors' information Qingyu Zhang is a Professor and Director of the Research Institute of Business Analytics and Supply Chain Management at Shenzhen University, China. He is a Distinguished Professor of Pearl River Scholar. Prior to this, he was a tenured professor at Arkansas State University, USA. He earned his PhD in Manufacturing Management and Engineering from the University of Toledo, USA, and M.E., B.E. and B.A. from Tsinghua University, China. His work was featured in JOM, EJOR, IJPR, IJPE, IJOPM, SCM, COR, CIE, JCP, and IEEE Transactions. He won the Jack Meredith Best Paper Award in the Journal of Operations Management in 2016. He serves on the Editorial Review Boards of Journal of Operations Management and Transportation Research Part E: Transportation and Logistics Review. Shimiao Yang is a Master's student of Management Science at Shenzhen University.

Availability of data and materials All data generated or analyzed during this study are included in this published article.

Authors' contributions Conceptualization (Q.Z., S.Y.); data curation (S.Y.); formal analysis (Q.Z., S.Y.); funding acquisition (Q.Z.); investigation (Q.Z., S.Y.); methodology (Q.Z., S.Y.); project administration (Q.Z.); resources (Q.Z.); software (Q.Z., S.Y.); supervision (Q.Z.); validation and visualization (Q.Z., S.Y.); writing—original draft (Q.Z., S.Y.); writing—review and editing (Q.Z., S.Y.). All authors read and approved the final manuscript.

Funding This research was funded by National Natural Science Foundation of China (#71572115); Guangdong 13th-Five-Year-Plan Philosophical and Social Science Fund (#GD20CGL28); Major Program of Social Science Foundation of Guangdong (#2016WZDXM005); and Natural Science Foundation of SZU (#836).

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

Ethics approval and consent to participate We got informed consent and approval from participants.

Consent for publication Not applicable.

References

- Addo-Tenkorang R, Helo PT (2016) Big data applications in operations/supply-chain management: a literature review. *Comput Ind Eng* 101:528–543
- Ahmad BW, SMK Q (2015) Big data promises value: is hardware technology taken onboard? *Ind Manag Data Syst* 115:1577–1595
- Arora S, Bala A (2020) A survey: ICT enabled energy efficiency techniques for big data applications. *Clust Comput* 23(2):775–796
- Boru D, Kliazovich D, Granelli F, Bouvry P, Zomaya AY (2015) Energy-efficient data replication in cloud computing datacenters. *Clust Comput* 18:385–402
- Brady GA, Kapur N, Summers JL, Thompson HM (2013) A case study and critical assessment in calculating power usage effectiveness for a data centre. *Energy Convers Manag* 76:155–161
- Capozzoli A, Primiceri G (2015) Cooling Systems in Data Centers: state of art and emerging technologies. *Energy Procedia* 83:484–493
- Chen C-T (2000) Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets Syst* 114:1–9
- Chen M, Mao S, Liu Y (2014) Big data: a survey. *Mobile Netw Appl* 19:171–209
- Cheng S, Zhang Q, Qin Q (2017) Big data analytics with swarm intelligence. *Ind Manag Data Syst* 116(4):646–666
- Cohen M (2018) Big data and service operations. *Prod Oper Manag* 27(9):1709–1723
- Corbett CJ (2018) How sustainable is big data? *Prod Oper Manag* 27:1685–1695
- Daim TU, Bhatla A, Mansour M (2013) Site selection for a data centre—a multi-criteria decision-making model. *Int J Sustain Eng* 6:10–22
- Dargie, W., Strunk, A. & Schill, A. Energy-aware service execution 2011 IEEE 36th conference on local computer networks, 4–7 Oct. 2011. 1064–1071
- Dimitrov DV (2016) Medical internet of things and big data in healthcare. *Healthc Inform Res* 22:156–163
- Fiandrino C, Kliazovich D, Bouvry P, Zomaya AY (2017) Performance and energy efficiency metrics for communication Systems of Cloud Computing Data Centers. *IEEE Trans Cloud Comp* 5:738–750

- Fisher M, Raman A (2018) Using data and big data in retailing. *Prod Oper Manag* 27:1665–1669
- Guitart J (2017) Toward sustainable data centers: a comprehensive energy management strategy. *Computing* 99:597–615
- Gumus AT (2009) Evaluation of hazardous waste transportation firms by using a two step fuzzy-AHP and TOPSIS methodology. *Expert Syst Appl* 36:4067–4074
- Ham S-W, Kim M-H, Choi B-N, Jeong J-W (2015) Simplified server model to simulate data center cooling energy consumption. *Energy and Buildings* 86:328–339
- Hammadi A, Mhamdi L (2014) A survey on architectures and energy efficiency in data center networks. *Comput Commun* 40:1–21
- Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Ullah Khan S (2015) The rise of “big data” on cloud computing: review and open research issues. *Inf Syst* 47:98–115
- Haywood A, Sherbeck J, Phelan P, Varsamopoulos G, Gupta SKS (2012) Thermodynamic feasibility of harvesting data center waste heat to drive an absorption chiller. *Energy Convers Manag* 58:26–34
- Herman J, Herman H, Mathews M, Vosloo J (2018) Using big data for insights into sustainable energy consumption in industrial and mining sectors. *J Clean Prod* 197:1352–1364
- Johnson H (2015). *Digging up dark data: What puts IBM at the forefront of insight economy* [online]. Available: <https://siliconangle.com/2015/10/30/ibm-is-at-the-forefront-of-insight-economy-ibminsight/> [Accessed July 17 2019]
- Jones GF, Fleischer AS (2014). A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery ... and the corresponding low-grade waste heat recovery opportunities. *Renew Sust Energ Rev* 31:622–638
- Kacfeh Emami C, Cullot N, Nicolle C (2015) Understandable big data: a survey. *Computer Science Review* 17:70–81
- Karnama A, Haghighi E, Vinuesa R (2019) Organic data centers: a sustainable solution for computing facilities. *Results in Engineering* 4: 1–3
- Kaya T, Kahraman C (2011) Multicriteria decision making in energy planning using a modified fuzzy TOPSIS methodology. *Expert Syst Appl* 38:6577–6585
- Kheybari S, Davoodi Monfared M, Farazmand H, Rezaei J (2020) Sustainable location selection of data centers: developing a multicriteria set-covering decision-making methodology. *Int J Inf Technol Decis Mak* 19(3):741–773
- Khuntia J, Saldanha T, Mithas S, Sambamurthy V (2018) Information technology and sustainability: evidence from an emerging economy. *Prod Oper Manag* 27(4):756–773
- Kim J-Y, Chang H-J, Jung Y-H, Cho K-M, Augenbroe G (2017) Energy conservation effects of a multi-stage outdoor air enabled cooling system in a data center. *Energy and Buildings* 138:257–270
- Krohling R, Campanharo V (2011) Fuzzy TOPSIS for group decision making: A case study for accidents with oil spill in the sea. *Expert Syst Appl* 38(4):4190–4197
- Kunkel JM, Shoukourian H, Heidari MR, Wilde T (2019) Interference of billing and scheduling strategies for energy and cost savings in modern data centers. *Sustain Comp: Informatics and Systems* 23: 49–66
- Kuo M-S, Tzeng G-H, Huang W-C (2007) Group decision-making based on concepts of ideal and anti-ideal points in a fuzzy environment. *Math Comput Model* 45:324–339
- Kwon S (2020) Ensuring renewable energy utilization with quality of service guarantee for energy-efficient data center operations. *Appl Energy* 276:1–11
- Lan S (2016) A customer satisfaction evaluation model for logistics services using fuzzy analytic hierarchy process. *Ind Manag Data Syst* 116:1024–1042
- Laney D (2001). 3D data management: controlling data volume, velocity, and variety. *Application Delivery Strategies*
- Lee AHI, Kang H-Y, Hsu C-F, Hung H-C (2009) A green supplier selection model for high-tech industry. *Expert Syst Appl* 36:7917–7927
- Lei H, Wang R, Zhang T, Liu Y, Zha Y (2016) A multi-objective co-evolutionary algorithm for energy-efficient scheduling on a green data center. *Comput Oper Res* 75:103–117
- Li L, Zheng W, Wang X, Wang X (2016) Data center power minimization with placement optimization of liquid-cooled servers and free air cooling. *Sustain Comp: Informatics and Systems* 11:3–15
- Lin C (2018) Profit optimization of sustainable low-to-medium temperature waste heat recovering management. *Ind Manag Data Syst* 118: 330–348
- Lucivero F (2020) Big data, big waste? A reflection on the environmental sustainability of big data initiatives. *Sci Eng Ethics* 26(2):1009–1030
- McKinsey. 2016. *Big data: Getting a better read on performance* [Online]. Available: <https://www.mckinsey.com/industries/high-tech/our-insights/big-data-getting-a-better-read-on-performance> [Accessed July 15 2019]
- Mehrjerdi YZ (2012) Developing fuzzy TOPSIS method based on interval valued fuzzy sets. *Int J Comput Appl* 42:7–18
- Melville N, Saldanha D, Rush D (2017) Systems enabling low-carbon operations: the salience of accuracy. *J Clean Prod* 166:1074–1083
- Möbius C, Dargie W, Schill A (2014) Power consumption estimation models for processors, virtual machines, and servers. *IEEE Transact Parallel Distrib Syst* 25:1600–1614
- Nadjahi C, Louahlia H, Lemasson S (2018) A review of thermal management and innovative cooling strategies for data center. *Sustain Comp: Informatics and Systems* 19:14–28
- Ni J, Bai X (2017) A review of air conditioning energy performance in data centers. *Renew Sust Energ Rev* 67:625–640
- Önüt S, Soner S (2008) Transshipment site selection using the AHP and TOPSIS approaches under fuzzy environment. *Waste Manag* 28: 1552–1559
- Oró E, Depoorter V, Garcia A, Salom J (2015) Energy efficiency and renewable energy integration in data centres. *Strategies and modelling review. Renew Sust Energ Rev* 42:429–445
- Oussous A, Benjelloun F-Z, Ait Lahcen A, Belfkih S (2018) Big data technologies: a survey. *J King Saud Univ – Comp Inform Sci* 30: 431–448
- Priyadumkol J, Kittichaikam C (2014) Application of the combined air-conditioning systems for energy conservation in data center. *Energy Build* 68:580–586
- Rahmani R, Moser I, Cricenti A (2020) Modelling and optimisation of microgrid configuration for green data centres: a metaheuristic approach. *Futur Gener Comput Syst* 108:742–750
- Ramli SJ, Jambari DI, Mokhtar UA (2017). A Framework Design for Sustainability of Green Data Center. *6th International Conference on Electrical Engineering and Informatics (ICEEI)*
- Rong H, Zhang H, Xiao S, Li C, Hu C (2016) Optimizing energy consumption for data centers. *Renew Sust Energ Rev* 58:674–691
- Saaty TL (1980) *The Analytic Hierarchy Process*. McGraw Hill International, New York
- Saaty TL & Vargas LG (2006). *Decision making with the analytic network process*, Springer US
- Sachs J (2012) From millennium development goals to sustainable development goals. *Lancet* 379:2206–2211
- Sanders NR, Ganeshan R (2018) Big data in supply chain management. *Prod Oper Manag* 27:1745–1748
- Seyhan S, Mehpare T (2010) The analytic hierarchy process and analytic network process: an overview of applications. *Manag Decis* 48:775–808
- Shaw K, Shankar R, Yadav SS, Thakur LS (2012) Supplier selection using fuzzy AHP and fuzzy multi-objective linear programming for developing low carbon supply chain. *Expert Syst Appl* 39: 8182–8192

- Sheme E, Holmbacka S, Lafond S, Lučanin D, Frashëri N (2018) Feasibility of using renewable energy to supply data centers in 60° north latitude. *Sustain Comp: Informatics and Systems* 17:96–106
- Shuja J, Gani A, Shamshirband S, Ahmad RW, Bilal K (2016) Sustainable cloud data centers: a survey of enabling techniques and technologies. *Renew Sust Energ Rev* 62:195–214
- Singh S, El-Kassar A (2019) Role of big data analytics in developing sustainable capabilities. *J Clean Prod* 213:1264–1273
- Song M, Wang S (2017) Participation in global value chain and green technology progress: evidence from big data of Chinese enterprises. *Environ Sci Poll Res* 24:1648–1661
- Sun C-C (2010) A performance evaluation model by integrating fuzzy AHP and fuzzy TOPSIS methods. *Expert Syst Appl* 37:7745–7754
- Todorovic MS, Kim JT (2014) Data centre's energy efficiency optimization and greening—case study methodology and R&D needs. *Energy and Buildings* 85:564–578
- Uddin M, Rahman AA (2012) Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics. *Renew Sust Energ Rev* 16:4078–4094
- Wang L, Zhang F, Aroca JA, Vasilakos AV, Zheng K, Hou C, Li D, Liu Z (2013) GreenDCN: a general framework for achieving energy efficiency in data center networks. *IEEE J Select Areas Commun* 32:4–15
- Webb M (2008) Smart 2020: enabling the low carbon economy in the information age. Climate Group 1:1–1
- Weihl B, Teetzel E, Clidas J, Malone C, Kava J, Ryan M (2011) Sustainable data centers. *XRDS: Crossroads, The ACM Magazine for Students* 17:8–12
- Whitehead B, Andrews D, Shah A, Maidment G (2014) Assessing the environmental impact of data centres part I: background, energy use and metrics. *Build Environ* 82:151–159
- Woodruff Z, Brenner P, Buccellato APC, Go DB (2014) Environmentally opportunistic computing: A distributed waste heat reutilization approach to energy-efficient buildings and data centers. *Energy and Build* 69:41–50
- Yu L, Chen L, Cai Z, Shen H, Liang Y & Pan Y 2018. Stochastic load balancing for virtual resource Management in Datacenters. *IEEE Transactions on Cloud Computing*, 1–1
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zhang H, Shao S, Xu H, Zou H, Tian C (2014) Free cooling of data centers: a review. *Renew Sust Energ Rev* 35:171–182
- Zhang P, Yan F, Du C (2015) A comprehensive analysis of energy management strategies for hybrid electric vehicles based on bibliometrics. *Renew Sust Energ Rev* 48:88–104
- Zhong RY, Newman ST, Huang GQ, Lan S (2016) Big data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives. *Comput Ind Eng* 101: 572–591

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.