OPEN ACCESS
EURASIA Journal of Mathematics Science and Technology Education
ISSN: 1305-8223 (online) 1305-8215 (print)
2017 13(8):4553-4589
DOI 10.12973/eurasia.2017.00951a

Disaster Recovery Site Evaluations and Selections for Information Systems of Academic Big Data

Chia-Lee Yang
National Center for High-Performance Computing, TAIWAN

Chi-Yo Huang
National Taiwan Normal University, TAIWAN

Yu-Sheng Kao
The University of Tokyo, JAPAN

Yi-Lang Tasi
National Cheng Kung University, TAIWAN

Received 1 October 2016 ▪ Revised 1 May 2017 ▪ Accepted 1 May 2017

ABSTRACT

The most dramatic factor shaping the future of higher education is Big Data and analytics. In the Big Data era, the explosive growth of massive data manipulations imposes a heavy burden on computation, storage, and communication in data centers. Increasing uncertainties in information system availability have become a daily serious problem. An appropriate evaluation and selection of the right information system disaster recovery (DR) site can ensure business continuity and investment optimization. However, most academic institutes always neglect the importance of DR. Not to mention the DR sites in the era of Big Data. Existing research results do not evaluate or select DR sites in general or those for academic Big Data applications in particular. Therefore, this research aims to establish an analytic framework for evaluating, selecting DR sites for academic Big Data. The proposed analytic framework is consisting of the Decision-Making Trial and Evaluation Laboratory (DEMATEL), DEMATEL-based network process (DNP) and VIšekriterijumsko KOmpromisno Rangiranje (VIKOR) methods. An empirical study based on a real Big Data DR application of an Asian high-performance computer center's evaluation and selection of DR sites for academic Big Data is used to illustrate the feasibility of the proposed framework. The analytic results can serve as a foundation for information technology administrators' strategies to reduce the performance gaps of a DR site for Big Data manipulations in general, and academic Big Data manipulations in special.

Keywords: big data, disaster recovery (DR), site selection, DEMATEL-based network process (DNP), VIšekriterijumsko KOmpromisno Rangiranje (VIKOR), multiple criteria decision making (MCDM).

INTRODUCTION

"Big Data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze (Manyika et al., 2011). The world of Big Data is poised to shake up everything from education, business, economics, the humanities, and every other aspect of society (Mayer-Schönberger & Cukier, 2014). Big Data is one of the two most dramatic factors shaping the future of higher education (Siemens & Long, 2011) and is

Correspondence: Chi-Yo Huang, *Department of Industrial Education, National Taiwan Normal University, Taiwan.*
georgeh168@gmail.com

State of the literature

- Although most of the national governments, the education systems, and the industry leaders have been notified about the importance of providing sufficient talents in Big Data related fields, the supply of the STEM talent pipeline does not meet the growing needs of our high-technology economy (Liebowitz, 2013).
- Sembiring and Siregar (2013) described the risk factors related to IT, IT organization, and business processes that influence DR site evaluation and selection from the business process perspective.
- Katal et al. (2013) pointed out three fundamental issue areas that need to be addressed to deal with Big Data: storage issues, management issues, and processing issues.

Contribution of this paper to the literature

- This research provides direct evaluation, selection, and improvement strategies for DR sites in general and the DR sites for Big Data applications in particular.
- This research defines an analytic framework for evaluating, selecting, and improving DR sites so as to reduce the gaps between the current status and the aspired level of all dimensions for selecting the best DR site.
- This research results provide IT managers with an understanding of the major concerns of DR site evaluation and selection for Big Data applications.

still a relatively new concept in educational research (Eynon, 2013; Zhang, 2014). Two areas under development oriented towards the inclusion and exploration of Big Data capabilities in education are educational data mining (EDM) and learning analytics (LA); most of recent researches in the two areas focus on science, technology, engineering, and mathematics (STEM) (Papamitsiou & Economides, 2014). Undoubtedly, analytics and Big Data have a significant role to play in the future of higher education (Siemens & Long, 2011).

The applications of academic Big Data analysis can be classified into two categories: improvement of the education service and development of students' competences of Big Data analytics, where analytics can be thought as the practice of mining institutional data to produce actionable intelligence (Campbell, deBlois, & Oblinger, 2007). The classification is consistent with the Long and Siemens' classification of LA and academic analytics (Siemens & Long, 2011). Here, LA is the measurement, collection, analysis and reporting of data about learners for understanding and optimizing learning and the environments (Ferguson, 2012; Siemens & Long, 2011). Academic analytics, in contrast, is the application of business intelligence in education and emphasizes analytics at institutional, regional, and international levels (Siemens & Long, 2011). Most of the recent researches on LA and education data mining focus on STEM (Papamitsiou & Economides, 2014). The development of students' competences of Big Data analytics can cross the gap between the supply of STEM talent pipeline versus the growing needs in the high-technology economy (Liebowitz, 2013).

The explosive growth of Big Data, characterized by data volume, variety, velocity, variability, veracity, and complexity, imposes a heavy burden on computation, storage, and communication in data centers (Boyd & Crawford, 2012). Existing information infrastructure may already be full of uncertainties; the same problems still exist, but are bigger with Big Data (Bahrami & Singhal, 2015). The burden from Big Data increases the uncertainties in the information system and generates business risks and data losses. In the extreme cases, the system crashes can cause business services disasters. Also, because the data sets need to be collected from both within and outside the organizations, securing sensitive data, protecting private information, and managing data quality lead to more challenges for information system availability. Thus, modern organizations are required to implement a long-term strategy and a complete solution, including a data back and disaster recovery (DR) plan for Big Data to enable business continuity (Chang, 2015).

An information system DR site is a second location which is physically isolated so that failures at one site are unlikely to propagate to the other (Sembiring & Siregar, 2013). Choosing the right remote DR site can ensures optimize investments for Big Data. Most information technology professionals and business stakeholders find DR site evaluation and selection is a tradeoff decision problem involved in justifying business risk of losses and additional operation costs (Alhazmi & Malaiya, 2012).

Apparently, information system DR sites are very critical for Big Data in general, and academic Big Data in special. Unfortunately, for years many academic institutions have ignored the importance of disaster management and continuity planning (Beggan, 2011; Kiernan, 2005a; Kiernan, 2005b). Not to mention the DR sites for the academic Big Data. Most existing research has focused on DR information technology with regard to Cloud Computing, storage technologies, or networking (Bahrami & Singhal, 2015; Chang, 2015; Hashem et al., 2015; Sengupta & Annervaz, 2014). Some research (e.g., Sahebjamnia, Torabi, & Mansouri, 2015; Thejendra, 2014) has introduced the construction of DR sites from the aspect of business continuity. In practice, solving DR site selection problems in the era of Big Data usually requires big investment and deviates from the overall welfare of the whole system (Beggan, 2011; Kiernan, 2005a; Kiernan, 2005b). However, very few past works have systematically analyzed factors for evaluating and selecting DR sites, the inter-relationships between the factors, or the weights associated with the factors. Furthermore, based on the authors' very limited knowledge, no prior work has tried to evaluate DR sites for academic Big Data applications. Thus, there is an urgent need to define a decision-making framework which can cover major aspects and factors for resolving real-world DR site evaluation problems in the Big Data era, while considering the inter-relationships between the aspects as well as various factors. Such a framework can help management in making decisions on evaluating, selecting, and enhancing DR sites for Big Data applications.

This research aims to define an analytic framework for evaluating and selecting DR sites so as to reduce the gaps between the current status and the aspired level of all aspects of selecting the best DR site. The analytic framework consists of the multiple criteria decision-making (MCDM) methods, which include the Decision-Making Trial and Evaluation Laboratory (DEMATEL), the DEMATEL-based network process (DNP), and VIšekriterijumsko KOmpromisno Rangiranje (VIKOR) methods. The DEMATEL is used to construct the network relation map (NRM) between factors (Fontela & Gabus, 1976). The DNP is used to derive the influence weights based on the concept of the Markov chain, which is introduced in the analytic network process (ANP). Then, the VIKOR method is used to integrate the performance gaps derived from all aspects and factors. An empirical case on evaluation and selection of DR sites for a real Big Data application for an Asian high-performance computer center, the institute which is in charge of most academic science and technology Big Data manipulations, is used to illustrate the feasibility of the proposed framework.

LITERATURE REVIEW

To review the current state of the art regarding DR site evaluation and selection strategy in the era of Big Data and construct an analytic framework accordingly, related literature is reviewed and summarized below. The literature focuses on academic Big Data, DR sites, and Big Data.

Academic Big Data and Science, Technology, Engineering and Mathematics Education (STEM)

Attempts to imagine the future of education often emphasize new technologies—ubiquitous computing devices, flexible classroom designs, and innovative visual displays; but the most dramatic factor shaping the future of higher education is Big Data and analytics. In the field of education, Big Data is still a relatively niche topic, but it is clearly beginning to grow (Eynon, 2013). Two areas under development oriented towards the inclusion and exploration of "Big Data" capabilities in education are EDM and LA (Papamitsiou & Economides, 2014). The areas of EDM and LA are both developing a number of identifiers characteristic of an established field of study (Romero & Ventura, 2010); and governments are beginning to produce reports on the potential of Big Data for education (Eynon, 2013).

The applications of academic Big Data analysis can be classified into two categories: improvement of the education service and development of students' competences of Big Data analytics, where analytics can be thought as the practice of mining institutional data to produce actionable intelligence (Campbell et al., 2007). These two applications of academic Big Data analysis are introduced further below. (1) Improvement of the education services:

The educational Big Data analysis is able to discover useful knowledge or interesting patterns from the unique type of data coming from educational settings (Baker & Yacef, 2009; Lam et al., 2015; Rabbany, Takaffoli, & Zaïane, 2011; Romero & Ventura, 2010). EDM looks for new patterns in data and develops new algorithms and/or new models, while LA applies known predictive models in instructional systems (Bienkowski et al., 2012). Higher education institutions are applying LA to improve visible and measurable targets such as grades and retention. The LA, in contrast, can be defined as the "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Ferguson, 2012; Siemens & Long, 2011). LA in higher education is used to predict student success by examining how and what students learn and how success is supported by academic programs and institutions (Mattingly, Rice, & Berge, 2012). K–12 schools and school districts are also starting to adopt such institution-level analyses for detecting areas for improvement, setting policies, and measuring results (Bienkowski et al., 2012). Besides, LA serves as a tool for closing the assessment loop in higher education (Mattingly et al., 2012). Recently, researchers and developers from the educational community started exploring the potential adoption of analogous techniques for gaining insight into online learners' activities (Papamitsiou & Economides, 2014). (2) Development of students' competences of Big Data analytics: The student's competences of Big Data analytics can be defined as the LA, which centers on the learning process (which includes analyzing the relationship between learner, content, institution, and educator). LA also help students develop sophisticated ways of working with data, comparable to what they will encounter in higher levels of education, at work, or elsewhere in their adult lives (Lee, 2013).

While the use of data science has become well established in STEM, the application of data science to education needs substantial research and development (CRA, 2015). Nowadays, more projects encourage students to get real-world data-sets and to solve complex problems. These Big Data programs have been developed in close partnership with information industry leaders such as IBM, Oracle, and SAS. Upon graduation, these data-savvy thinkers are in high demand and ready to make an impact. Although most of the national governments, the education systems, and the industry leaders have been notified about the importance of providing sufficient talents in Big Data related fields, the supply of the STEM talent pipeline does not meet the growing needs of our high-technology economy. Much of the promise of Big Data analytics is contingent on ample and a growing supply of STEM talent (Liebowitz, 2013). In summary, not only will there be a substantial increase in demand for people with the skills required to allow our economy to take advantage of this technology, but also that supply, given the momentum view, is not increasing and will face increased international competition for people with these skills across the STEM fields (Liebowitz, 2013).

Information system Disaster Recovery site problems

As information systems have become universal and powerful tools, organizations have come to expect their information systems to be available without interruption, even during disasters (Cerullo & Cerullo, 2004). However, unpredictable conditions or underestimated risks such as a natural disaster, a technical accident, or a system failure occur occasionally (Anthopoulos, Kostavara, & Pantouvakis, 2013). Recent disasters, such as data assaults, natural catastrophes, and terrorist attacks, have told us that disasters can hit organizations of every size, threatening to disrupt and potentially even destroy those that are not fully prepared (Michael Wallace & Webber, 2010). Disasters directly interrupt information systems operation, disaffect customers, and compromise business credibility and revenue streams (Gibb & Buchanan, 2006). Therefore, the establishment of a viable DR system for business continuity is important to mitigate risks (Han, Li, & Zhu, 2012), no matter from the aspects of business managers and academicians (Rodger, Bhatt, Chaudhary, Kline, & McCloy, 2015).

DR is a detailed, step-by-step action plan for quickly recovering after a natural or manmade disaster. The details of DR vary depending on the business needs and can be developed in-house or purchased as a service (TechAdvisory, 2010). A scan of the DR literature shows an increasing number of publications over the past decade (Mayer, Moss, & Dale, 2008; Sutton & Tierney, 2006; Tierney, 2007; Webb, Tierney, & Dahlhamer, 2000). Since 2001, three events (September 11 attacks, Hurricanes Katrina and Rita, and pandemic influenza planning) have each spurred a selection of articles focusing on DR from the computer science perspective in particular (Dawes, Cresswell, & Cahan, 2004; Sutton & Tierney, 2006). Information system DR is a set of activities executed once the disaster occurs, including the use of backup facilities to provide users of IT systems with access to data and functions required to sustain business processes (Wiboonrat, 2008). The risks of business interruption expand as companies become more dependent on information technology infrastructure. However, the infrastructure of today is not perfect; often it is the result of years of acquisition, investment, rebuilding, making partial improvements, and instituting temporary fixes that became permanent. Such an infrastructure significantly increases the risks to responses to disaster (Clitherow, Brookbanks, Clayton, & Spear, 2008). Information infrastructure availability has become a critical issue for DR (Yang, Yuan, & Huang, 2015). A comprehensive approach to business continuity planning seeks to mitigate all major interruptions of business systems with backup systems. Thus, the backup information infrastructure has become critical to enterprises (Han et al., 2012).

An information system DR site is a backup data center that aims to replicate data to a remote location and synchronize those data with the primary site in case of the primary site failure (Sembiring & Siregar, 2013). Information system DR sites can restore data and keep an organization's IT system operating during or after a disaster (Sembiring & Siregar, 2013). The field of information system DR studies has a small but increasing number of studies. Different aspects related to computer science and facility location have been considered. Some researchers have focused on DR technologies such as virtualization performance, network technology, and storage. For example, Sengupta and Annervaz (2014) provides a data distribution planner for DR as a schematic diagram of backing-up critical business data into data centers across several geographic locations. Some studies explore DR site problems form disaster management science viewpoint. For example, Sembiring and Siregar (2013) described the risk factors related to IT, IT organization, and business processes that influence DR site evaluation and selection from the business process perspective. Recently, Sahebjamnia et al. (2015) integrated business continuity and DR planning for efficient and effective resuming and recovering of critical operations after being disrupted.

Some studies have been based on a physical location scenario. For example, Ablanedo-Rosas, Gao, Alidaee, and Teng (2009) allocated emergency/recovery centers throughout the state in such a way that all municipalities are within 55 kilometers of the closest center. Balcik and Beamon (2008) proposed a model to determine the number and locations of distribution centers in a relief network by setting service level requirements. Recently, Al-Shaikh, Al-Hussain, and Al-Sharidah (2015) tried to select an IT Disaster Recovery Site for Oil and Gas Companies from 6 candidate facilities within the range of a distance of 65 to 425 kilometers. Above are practical examples for information system DR site selection researches during the past decade.

Big Data and Disaster Recovery

The development and maturity of cloud computing technologies has enabled exponential growth in the data that are produced, processed, stored, shared, analyzed, and visualized (Garlasu et al., 2013; Tian & Zhao, 2015). Data generation has increased drastically over the past few years, leading enterprises dealing with data management to swim in an enormous pool of data. The world's volume of data doubles every 18 months, for example, and enterprise data are predicted to increase by about 650% over the next few years (Beyer & Laney, 2012). Today, most firms have more data than they can handle and managers recognize the potential for value, but the promise of Big Data still has not been realized, according to leading academic and business media sources (Chang, Kauffman, & Kwon, 2014).

Although in ubiquitous use today, the term of Big Data as a concept is nascent and has uncertain origins (Gandomi & Haider, 2015). Big Data definitions have evolved rapidly. Some scholars have defined Big Data from an information technology viewpoint. For example, Zikopoulos, Eaton, DeRoos, Deutsch, and Lapis (2012) defined Big Data as the amount of data beyond existing information technology's capability to store, manage, and process efficiently. Hashem et al. (2015) suggested that Big Data result in an increasing volume of data that are difficult to store, process, and analyze through traditional database technologies. Meanwhile, another group of researchers have defined Big Data from a data character perspective. For example, Berman (2013) defined Big Data as being characterized by three Vs: volume, variety, and velocity. Gartner defined Big Data as high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information process for enhanced insight and decision making (Beyer & Laney, 2012). "Volume" refers to large amounts of data. "Variety" refers to data that come in different forms, including traditional databases, images, documents, and complex records. “Velocity” refers to the content of the data, which is constantly changing through the absorption of complementary data collections, through the introduction of previously archived data or legacy collections, and from streamed data arriving from multiple sources. The three Vs (3Vs) have emerged as a common framework to describe Big Data (Zikopoulos et al., 2012).

In addition to the 3Vs, the consulting company IDC specified that Big Data is not only characterized by the three Vs mentioned above but also by a fourth V-value (IDC, 2009). IDC defined Big Data as “a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high velocity capture, discovery, and/or analysis.” This 4V definition is widely recognized because it highlights the meaning and necessity of Big Data (Gantz & Reinsel, 2011). Furthermore, the SAS company introduced “Variability” and “Complexity” as two additional dimensions of Big Data. Variability refers to the variation in the data flow rates. Often, Big Data velocity is not consistent and has periodic peaks and troughs. Complexity refers to the fact that Big Data are generated through a myriad of sources. This imposes a critical challenge: the need to connect, match, cleanse, and transform data received from different sources. To sum up, this research defines Big Data as the amount of data beyond technology’s capability to store, manage, and process efficiently.

Big Data have amassed significant attention from researchers in information sciences and policy and decision makers in governments and enterprises (Chang et al., 2014; Chen & Zhang, 2014; Lin, Shuang, Yifang, & Shouyang, 2014). Scholars including computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, and sociologists are clamoring for access to the massive quantities of Big Data produced by and about people, things, and their interactions (Boyd & Crawford, 2012). Research on Big Data related to computer science has focused on system performance and scalability technology, such as virtualization, Hadoop, MapReduce, and security for increases in heterogeneous environments (Andreolini, Colajanni, Pietri, & Tosi, 2015; Kshetri, 2014). Few studies have focused on Big Data for system availability, DR in particular (Clitherow et al., 2008; Serrelis & Alexandris, 2006). However, more organizations now view Big Data as a strategic asset so data must be well preserved to enable reuse. Because digital data are so easily shared, replicated, and recombined, the costs of Big Data backup and recovery are high. Organizations are demanding that researchers and host institutions implement data management plans that address the full life cycle of data, including what happens after a disaster, and are seeking the greatest payoff for their investments (Yang, Li, & Yuan, 2014).

Significant questions for Big Data backup and recovery are emerging. Because Big Data volumes exceed the capacity of current storage and processing systems, Big Data represent a large set of technical research problems in Big Data DR naturally. Katal et al. (2013) pointed out three fundamental issue areas that need to be addressed to deal with Big Data: storage issues, management issues, and processing issues. In particular, the storage available is not sufficient for storing the large amount of data which are being produced by almost everything. Uploading these large amounts of data to the cloud may be an option. Lin et al. (2014) discussed the big network capacity challenge involves storing and recovering data in one DR location. Terabytes of data takes large amounts of time to be uploaded to a DR site. Minelli, Chambers, & Dhiraj (2003) discussed that the value of Big Data is from data scientist’s analyze. Because the researchers are hard to guarantee the analyst result is valuable. IT professionals are difficult to evaluate the critical level of Big Data. Kshetri (2014) discussed that the purpose of Big Data is a potential goldmine for cybercriminals, which leads to amplified information security technical and facility costs. A higher data set volume increases the probability that the data files and documents may contain inherently valuable and sensitive information; it also makes a more appealing target for hackers.

Consideration factors of Disaster Recovery site evaluation and selection

Many organizations implement a business continuity management plan to avoid turning away customers, configuration resilience, or obligation (Herbane, Elliott, & Swartz, 2004). Regulators and international standards are the most important concerns for organizations (Tammineedi, 2010). For example, the U.S. National Institute of Standards and Technology (NIST) enacted regulations of the SP 800-34 in 2006, which require high- and medium-level public sectors to set up DR sites (Bowen, Hash, & Wilson, 2006). International standards related to information system DR include those from the International Standard Organization (ISO 22301, ISO 22313), International

Electro-technical Commission (IEC 27001, 27002, 27031), and International Telecommunication Union (ITU -TL.92, ITU-T L.1300) (Yang et al., 2015). These regulations and international standards lead to a number of drivers and requirements to minimize the exposure to IT operational risk, improve IT operational resilience, and support the goal of continuously available IT service, as applied to the infrastructure.

In addition to the above, prior research has discussed the consideration factors of DR sites from different perspectives. Cegiela (2006) explored the processes of preparation for DR, such as organizational structure and technical facilities, as well as considerations associated with business requirements. The author also discussed characteristics of technological solutions for DR sites such as time parameters, practical availability, coverage, and relative costs. Wiboonrat (2008) investigated fundamental requirements of banking business units for mapping criticality of business continuity to DR readiness and guidelines for creating banking standardized procedures for DR plans. Clitherow et al. (2008) considered the DR site not as geographically separated, but as having spare capacity to ensure that all required applications can operate in either of the facilities in the event of the loss of the other facility, replication of data between the data centers, high levels of physical and network security, and a safe system for storing data. Yang et al. (2015) examined the consideration factors of DR site selection from five perspectives: location and infrastructure, IT facility, DR objectives, DR readiness exercises, and operation management.

Based on the literature review, most scholars agree that site selection decision-making frameworks vary for different problems and no site selection plans have been proposed to deal with unpredicted phenomena (Awasthi, Chauhan & Goyal, 2011). Although most real-world organizations have established DR sites based on international standards and national regulations requirement, international standards are sometimes very high level and abstract. Organizations need to go detail as operation guideline. Few studies have carried out on the selection of DR sites using the hybrid MCDM model combining with DANP and VIKOR. To date, only Yang et al. (2015) have studied the consideration factors concerning DR site selection for data centers. Therefore, this research adopts the consideration factors guideline from Yang et al. (2015) and uses a real Big Data DR application of an Asian high-performance computer center's evaluation and selection of DR sites to illustrate the feasibility of the proposed framework. The definitions of consideration factors are shown in **Table 1**. The corresponding symbols are defined also. This study proposed an analytical framework by using DEMATEL combined with ANP to determine the degrees of influence among the factors and leveraging VIKOR method for calculating the compromise ranking and gap of the alternatives for alternative improvement.

METHODOLOGY

In this Section, a hybrid MCDM framework consisting of the DEMATEL, DNP and VIKOR method is developed for DR site selections. First, the DEMATEL (Fontela & Gabus, 1976) technique is used to construct an network relations map (NRM) and derive the interrelations between factors. Then, the DEMATEL based Network Process (DNP) is used to derive the influence weights based on the basic concepts of the Analytic Network Process (ANP) being proposed by Saaty (1999). The VIKOR method presented by Opricovic (1998) with the influence weights being derived by the DNP is used to integrate the performance gaps from factors to aspects. The flowchart of the decision-making framework is illustrated in **Figure 1**. This hybrid MCDM model have been used to resolve various selection problems such as brand evaluation (Wang & Tzeng, 2012), smartphone technology acceptance (Huang & Kao, 2015), materials for engineering designs (Liu, You, Zhen, & Fan, 2014), the service selection in three dimensional printing industry (Liao, Wu, Huang, Kao, & Lee, 2014), low carbon suppliers (C. W. Hsu, Kuo, Shyu, & Chen, 2014), and RFID technology evaluation and selection (Lu, Lin, & Tzeng, 2013).

Table 1. Candidate Aspects and Factors for Evaluating DR Sites

Aspects	Criteria	Descriptions
Location and Infrastructure (A)	Natural Disaster (a_1) (Snedaker, 2013)	Appropriate DR site should minimize influences by natural disasters. The natural hazards include earthquakes, flood, hurricanes, typhoon and geological hazards.
	Manmade Disaster (a_2) (Snedaker, 2013)	Appropriate geographic location prevents the human-caused hazards. The manmade disasters include terrorism, bomb, explosion, fire, cyber-attack, civil disorder, protests, product tampering, radioactive contamination, embezzlement, kidnapping, extortion, and subsidence.
	Distance From Primary Site (a_3) (Wallace, Webber, & Webber, 2011)	The distance between the primary and backup recovery sites depends on the risk assessment; the recovery site must be far enough away so that the same catastrophe does not strike both sites
	Transportation (a_4) (Chang et al., 2012; Hanaoka & Qadir, 2005)	The transport network including roads, airports, port, and railways. The transport system is critical during a natural disaster due to its pivotal role in resourcing recovery. The high cost of resource transportation and lack of transport alternatives were major barriers to post-disaster reconstruction.
	Electricity and Cooling (a_5) (Gregory, 2011; Wallace et al., 2011)	Appropriate DR site should have stable power and a cooling system to prevent power outages and system shut down. The major consideration of electricity include: monitor the line and filter out spikes, provision of additional power, and ensure the transition from normal power supply to emergency power supply without loss of service to critical devices. Meanwhile, as the amount of heat that newer equipment discharges per square foot of space, cooling equipment are becoming daily important.
	Detection and Monitoring (a_6)(Whitson, 2003)	The DR site's building should have fast detection, monitoring alarm and operate equipment and design. The management of a computer security system involves intrusion detection and monitoring of the entire enterprise's computers.
IT System Availability (B)	Backup Strategies (b_1) (Wallace et al., 2011)	Proper backups of critical data can survive the organization from a disaster. Effective backups that completely protect critical data require thorough planning. The backup strategies include full system backup, incremental backup, and differential backup.
	Backup Servers (b_2) (Khoshkholghi et al.,2014)	Back up in physical backup servers, including local or geographical redundancy servers. Any backup solution must maintain the transactional integrity of the data so that, when the data is restored.
	Backup System Architecture (b_3) (Brooks, Bedernjak, Juran, & Merryman, 2002)	Backup system architecture and planning plays a critical role in the recovery phase as systems are rebuilt, applications are restored, data is recovered, and systems are put back online into production. At the completion of the recovery phase, systems will be functioning to the extent determined in the plan. Beyond the critical recovery phases, less critical recovery operations may ensue.
	Telecommunication Infrastructure (b_4) (Wallace et al., 2011)	The infrastructure of telecommunications, which includes internet bandwidth, fiber backbone route, and transaction time/latency. Key concerns include natural and man-made hazards, telephone equipment room (temperature, humidity, etc.), internal and external cabling, and route separation.
	Carrier and Support (b_5) (Rothstein, 2007)	Telephone companies and long distance carriers offer a wide range of virtual network services; loss of virtual network services, like traditional long distance service, can severely impair a company's ability to conduct business. All carriers being present in the vicinity and their support and service models in place. e.g., different source of carriers to avoid unexpected interruption by one carrier.

Table 1 (continued). Candidate Aspects and Factors for Evaluating DR Sites

Aspects	Criteria	Descriptions
DR Objectives (C)	Recovery Point Objective (c_1) (Roebuck, 2012)	Recovery Point Objective describes the acceptable amount of data loss measured in time; the Recovery Point Objective is the point in time to which data must be recovered as defined by the organization.
	Recovery Time Objective (c_2) (Roebuck, 2012).	The Recover Time Objective is the duration of time and a service level within which a business process must be restored after a disaster (or disruption) in order to avoid unacceptable consequences associated with a break in business continuity
	Testing and Exercises (c_3) (Roebuck, 2012)	The purpose of testing is to achieve organizational acceptance that the business continuity solution satisfies the organization's recovery requirements. Testing may include: crisis command team call-out testing, technical swing test from primary to secondary work locations, technical swing test from secondary to primary work locations, application test, and the business process test. Three types of exercise can be employed when testing the business continuity plan: simple, medium and complex exercises.
Disaster Readiness exercises (D)	Education and Training (d_1) (Smith, 2012)	Regular training and education programs for disaster recoveries, which include operations, technology, information security, and DRP processes. According to Smith (2012), training, education, and outreach initiatives; and the funding needed to implement them—can play an important but often overlooked role in recovery.
	DR Work Area (d_2) (Gregory, 2011)	Establish an area where critical IT workers can work during and after the recovery operation (Gregory, 2011). The physical precinct in a DR center, including operation area, equipment handling, and testing area.
	Emergency Operations Center (d_3) (Wallace et al., 2011)	An Emergency Operations Center is a physical place where all the communications of the recovery effort are focused. The Emergency Operations Center is essential when addressing serious or wide-scale disasters. It allows a company's management to reestablish organizational leadership, allocate resources, and focus on emergency containment and recovery.
Operation management (E)	Project Management (e_1) (Snedaker, 2013)	The project management plan of a DR site, such as the plan–do–check–act continuous quality improvement model. Elements of project success include: executive support, user involvement, experienced project manager, clearly defined project objectives, clearly defined project requirements, clearly defined scope, shorter schedule, multiple milestones, and clearly defined project management process.
	Information Security Management Procedure (e_2) (Snedaker, 2013)	A DR site should be certified by information Security standards and renewed regularly, such as ISMS, BS 7799, etc. The information security management should be based on an analysis of the most recent, more developed information security statutes, and responsibility for compliance increasingly rests with the BoD or CEO. The development and maintenance of a process-oriented written information security program (WISP) is critical to the ability of a company to meet its legal obligations as it relates to the management of information security.
	DR Procedure (e_3) (Judson, 2012)	Disaster management is the discipline of dealing with and avoiding risks; the availability of a good DR procedure and emergency response plan in place is essential for successful; DR. A DR site should have well-documented process of DR plan to recover and protect a business IT infrastructure in the event of a disaster.
	Top Manager's Supporting (e_4) (Snedaker, 2013)	Top manager's supports and commitment for DRP operations, including allocating time and resources required in the DR. Executive support for any IT project is typically the number one success factor.
	Resources (e_5) (Chang et al., 2012)	In DR projects, the operational environment is uncertain, complex, and dynamic. The "business as usual" way of managing resources may not be fully applicable. Evidence shows that post-DR projects are more likely to suffer resource shortages and supply disruption. These resourcing problems contribute to final recovery project failures.

Source: Adapted from Yang et al. (2015)

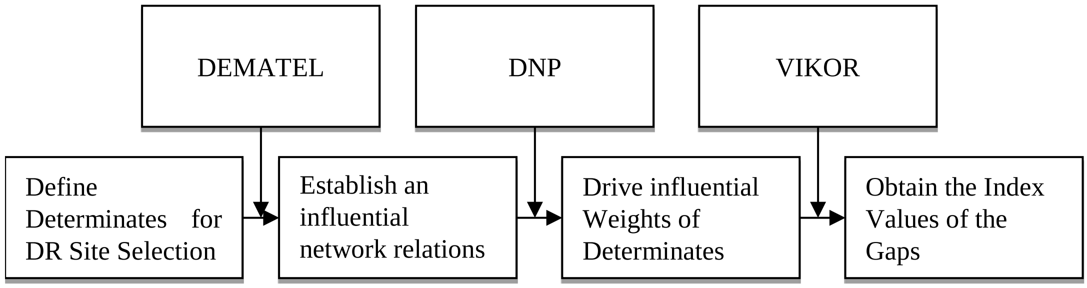

Figure 1. The flowchart of decision making framework

The Modified Delphi method

The Delphi method, designed by Dalkey and Helmer (1963), is a method for structuring a group communication process to facilitate group problem solving and to structure models. The method can be applied to problems that do not lend themselves to precise analytical techniques, but rather that could benefit from the subjective judgments of individuals on a collective basis and to focus their collective human intelligence on the problem at hand (Linstone & Turoff, 2002). The Delphi method is a mature and very adaptable research method used in many information systems and IT research arenas (Gallego & Bueno, 2014). Murry and Hammons (1995) modified the traditional Delphi technique by simplifying the step of conducting the first round of a survey and replaced the conventionally adopted open style survey. The modified Delphi technique is similar to the full Delphi in terms of procedure and intent. The advantages of the modified Delphi method include time savings and a focus on research themes, eliminating the need for speculation on the open questionnaire and an improvement in the response of the main topic (Liao et al., 2014). Accordingly, in this investigation we develop dimensions and criteria for evaluating the DR sites for information systems of academic Big Data by combining the modified Delphi method with interviews of anonymous experts.

Decision Making Trial and Evaluation Laboratory (DEMATEL)

The DEMATEL method was originated from the Geneva Research Centre of the Battelle Memorial Institute (Fontela & Gabus, 1976), can convert complex systems into a clear causal structure which simplifies the interrelationships among consideration factors. The DEMATEL method constructs the interrelations between factors/criteria to build a network relations map (NRM). The methodology can confirm interdependence among variables/criteria and restrict the relationships that reflect characteristics within an essential systemic and developmental trend. The DEMATEL technique has been successfully applied in many situations, such as identifying key successful factors in emergency management (Zhou, Huang, & Zhang, 2011), risk control assessment (OuYang, Shieh, & Tzeng, 2013), and risk factors of IT outsourcing (Fan, Suo, & Feng, 2012). As this research is to identify the causation and influence strengths of the consideration factors, this research employed the quantitative DEMATEL method in this research. The method can be summarized as follows

Step 2-1: Build an initial direct-relation matrix.

Experts are asked to indicate the direct influence degree of direct influence each factor i exerts on each factor j, as indicated by a_{ij}. The initial direct-relation matrix $\boldsymbol{A}$ is obtained by pairwise comparisons through equation (1). The influence degree a_{ij} represents the strength of influence or influenced relationships between the indented criteria. The higher influence degree a_{ij} represents that the experts express that factor i the stronger possible direct influence on the factor j. A five-point Likert scale is used to evaluate the influence degree ranging from 0 to 4, which indicates 'no influence (0)', 'very low influence (1)', 'low influence (2)' to 'high influence (3)', and 'very high influence (4)', respectively (Hsu, Kuo, Chen, & Hu, 2013; OYang et al., 2013). The influence degree is defined by the experts to perform pairwise comparisons of the factors and determine the causalities (Tzeng & Huang, 2012).

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

Step 2-2: Normalize the direct-relation matrix.

The normalized direct-relation matrix $\boldsymbol{N}$ is obtained through equation (2) and (3).

$$\boldsymbol{N} = y\boldsymbol{A} \tag{2}$$

$$y = \min\left\{1/\max_i \sum_{j=1}^{n} a_{ij}, 1/\max_j \sum_{i=1}^{n} a_{ij}\right\}, i, j \in \{1,2,...,n\} \tag{3}$$

Step 2-3: Attain a total relation matrix $\boldsymbol{T}$. The total-relation matrix $\boldsymbol{T}$ is acquired by equation (4), where $\boldsymbol{I}$ is the identity matrix.

$$\begin{aligned} \boldsymbol{T} &= \boldsymbol{N} + \boldsymbol{N}^2 + \boldsymbol{N}^3 + ... + \boldsymbol{N}^{\varepsilon} \\ &= \boldsymbol{N}\left(\boldsymbol{I} + \boldsymbol{N} + \boldsymbol{N}^2 + ... + \boldsymbol{N}^{\varepsilon-1}\right)\left(\boldsymbol{I} - \boldsymbol{N}\right)\left(\boldsymbol{I} - \boldsymbol{N}\right)^{-1} \\ &= \boldsymbol{N}\left(\boldsymbol{I} - \boldsymbol{N}^{\varepsilon}\right)\left(\boldsymbol{I} - \boldsymbol{N}\right)^{-1} \\ &= \boldsymbol{N}\left(\boldsymbol{I} - \boldsymbol{N}\right)^{-1}, \text{ when } \varepsilon \to \infty, \boldsymbol{N}^{\varepsilon} = [0]_{nx\,n} \end{aligned} \tag{4}$$

Step 2-4: Calculate the influence strength of the factors.

Aggregate the values of the rows and columns in matrix $\boldsymbol{T}$ to obtain r and c vectors through the equation (5) and (6) respectively.

$$\boldsymbol{T} = \left[t_{ij}\right], i, j \in \{1,2,....n\}$$

$$\boldsymbol{r} = [r_i]_{nx1} = \left(\sum_{j=1}^{n} t_{ij}\right)_{nx1} \tag{5}$$

$$\boldsymbol{c} = \left[c_j\right]_{nx1} = \left(\sum_{i=1}^{n} t_{ij}\right)'_{1xn} \tag{6}$$

The r_i is the row sum of the i[th] row of matrix $\boldsymbol{T}$. Thus, r_i represents the sum of the influences dispatching from factor i to the other factors. The c_j is the column sum of the j[th] column of matrix $\boldsymbol{T}$. Thus, c_j presents the influences that factor i is receiving from the other factors. When $i = j, r_i + c_j$ provides an index of the strength of influences given and received, that is, $r_i + c_j$ shows the degree of the central role that the factor i plays in the problem. In addition, the difference $r_i - c_j$ shows the net effect that factor i contribute to the problem. If $r_i - c_j$ is positive, then factor i is affecting other factors, and if $r_i - c_j$ is negative, then factor i is being influenced by other factors (Tzeng & Huang, 2012).

Step 2-5: Set a threshold value and obtain the NRM.

Since the map would be too complex to show all the network of determinates form matrix $\boldsymbol{T}$ converts to the NRM. It is necessary to sets a threshold value α for the influence level to filter out minor effects. The identification of the threshold value of DEMATEL is based on Lenth's principles of distinguishing effect significance, whereby threshold value is adopted to eliminate non-significant factors for obtaining factors with significant influences in scenarios with complex problems or factors (Hsieh, Lee, & Lin, 2016). When constructing the NRM, the threshold value is to filter out negligible effects while maintaining the complexity of the system as a whole to a manageable level (Huang, Huang, &Yang, 2015). The threshold value can be deduced by statistical, natural language or experts (Hsieh et al., 2016; Li & Tzeng, 2009; Tzeng & Huang, 2012). The influence values in matrix $\boldsymbol{T}$ which are higher than the threshold value will be chosen and converted into the NRM. When the threshold value and the relative NRM have been decided, the NRM can be drawn accordingly.

DEMATEL-based Analytic Network Process (DNP)

This study seeks to select a DR site, which usually consists of multiple dimensions and criteria and to determine the influential weights of those criteria. In a traditional methods, researchers use Analytic Network Process (ANP), which published by Saaty (1999)(Saaty, 1999), that can systematically overcome all types of dependences. The initial step is to compare the criteria in the entire system to form an unweighted supermatrix by pairwise comparisons in ANP procedures. Each criterion in a column is divided by the number of clusters, and thus assumes that each cluster has the same weight. However, each cluster may be different in degree and equal weight in obtaining the weighted supermatrix is not reasonable(Ou Yang, Shieh, Leu, & Tzeng, 2008). In a complex system, all criteria are directly or indirectly related. Consequently, the improved ANP (named DNP) which based on the basic concept of the ANP by using the DEMATEL technique can improve this shortcoming and obtain results. The DEMATEL technique is used to build an NRM for each criterion and dimension and also to improve the normalization process of the traditional ANP. DNP methodology can verify the interdependence of variables and attributes, building a relationship that reflects those characteristics with an essential system and evolutionary trend.

Because the DEMATEL method is based upon the graph theory, the NRM is a graph that convert the relationship between the causes and effects of criteria into an intelligible structural model of the system. NRM also portrays a contextual relationship between the elements with the strength of influence. This study uses influence levels values of each element over others from NRM as the basis of the normalization supermatrix for determining ANP weights to obtain the relative importance. The steps for building a NRM using the DEMATEL technique are summarized below.

Step 3-1: Find the normalized total influence matrix T_D^{nor}. The total-influential matrix T_D needs to be normalised by dividing it by the following formula.

$$t_D^i = \sum_{j=1}^{m} t_D^{ij} \tag{7}$$

$$T_D = \begin{bmatrix} t_D^{11} & \cdots & t_D^{1j} & \cdots & t_D^{1n} \\ \vdots & & \vdots & & \vdots \\ t_D^{i1} & \cdots & t_D^{ij} & \cdots & t_D^{\alpha in} \\ \vdots & & \vdots & & \vdots \\ t_D^{n1} & \cdots & t_D^{nj} & \cdots & t_D^{nn} \end{bmatrix} \tag{8}$$

$$T_D^{nor} = \begin{bmatrix} t_D^{nor11} & \cdots & t_D^{nor1j} & \cdots & t_D^{nor1n} \\ \vdots & & \vdots & & \vdots \\ t_D^{nori1} & \cdots & t_D^{norij} & \cdots & t_D^{norin} \\ \vdots & & \vdots & & \vdots \\ t_D^{norn1} & \cdots & t_D^{Fnornj} & \cdots & t_D^{nornn} \end{bmatrix} \tag{9}$$

Step 3-2: Build an unweighted supermatrix W_c. the total influential matrix is normalised into a supermatrix according to the interdependence between the relationships of the dimensions and clusters to obtain an unweighted supermatrix W_c as shown in Eq. (10).

$$W_c = (T_c^{nor})' = \begin{bmatrix} W^{11} & \cdots & W^{1j} & \cdots & W^{1n} \\ \vdots & & \vdots & & \vdots \\ W^{i1} & \cdots & W^{ij} & \cdots & W^{in} \\ \vdots & & \vdots & & \vdots \\ W^{n1} & \cdots & W^{nj} & \cdots & W^{nn} \end{bmatrix} \tag{10}$$

Step 3-3: Find the influential weights of the DNP. Using Eq 11, a weighted super-matrix W_c^* can be obtained by the product of T_D^{nor} and W_c. This result demonstrates that these influential level values are the basis of normalization to determine a weighted super-matrix.

$$W^*_c = T_D^{nor} W_c = \begin{bmatrix} t_D^{nor_{11}} W_c^{11} & \cdots & t_D^{nor_{i1}} W_c^{i1} & \cdots & t_D^{nor_{m1}} W_c^{m1} \\ \vdots & & \vdots & & \vdots \\ t_D^{nor_{1j}} W_c^{1j} & \cdots & t_D^{nor_{ij}} W_c^{ij} & \cdots & t_D^{nor_{mj}} W_c^{mj} \\ \vdots & & \vdots & & \vdots \\ t_D^{nor_{1m}} W_c^{1m} & \cdots & t_D^{nor_{im}} W_c^{im} & \cdots & t_D^{nor_{mm}} W_c^{mm} \end{bmatrix} \tag{11}$$

VIKOR method

The VIKOR method (Opricovic, 1998) is used to integrate the performance gaps from criteria to dimensions of complex systems. It determines the compromise ranking list and the compromise solution, and the weight stability intervals for the preferred stability of the compromise solution can be obtained from the initial weights given by the AHP or ANP in the traditional method (Opricovic & Tzeng, 2004). This method focuses on ranking and selection from a set of alternatives in cases of conflicting criteria. It introduces a multi-criteria ranking index based on the particular measure of "closeness" to the "ideal" solution (Opricovic, 1998). This study use VIKOR method to obtain the aspiration level and knowing how to improve and create improving strategies for DR site selection. Assuming that the feasible alternatives are represented by $A_1, A_2, .., A_m$ the rating/performance of the alternative A_i with respect to the criterion c_j is denoted as f_i^*; w_j is the weight of the j[th] criterion, where j=1, .., n, and $\boldsymbol{n}$ is the number of criteria. Then the compromise ranking algorithm of modified VIKOR includes the following steps based on Opricovic and Tzeng (2004) and Liao et al. (2014):

Step 4-1: Determine the positive-ideal level. The positive-ideal level can be determined using Eq. (12) as below. Different from the traditional VIKOR method: $f_i^* = {}^{max}_{\ j} f_{ij}$, $f_i^- = {}^{min}_{\ j} f_{ij}$ the performance definition given here can avoid ''choose the best among inferior options or alternatives'' and thus, is more appropriate for the selection of materials in the real-world.

f_i^*= Positive-ideal level, j=1, 2, …, n (12)

Step 4-2: Normalize the original rating matrix. The original performance matrix $[f_{ij}]_{mxn}$ can be converted into a normalized performance matrix using Eq. (13).

$$r_{ij} = (\left| f_i^* - f_{ij} \right|) / (\left| \max(f_i^*, f_i^{\max}) - \min(f_i^*, f_i^{\max}) \right|) \tag{13}$$

Step 4-3: Compute the values S_j and $R_j, j = 1, 2, .., m$ by the relations.

$$S_j = \sum_{i=1}^{n} w_i (f_i^* - f_{ij}) / (f_i^* - f_i^-),\ R_j = \max_i \left[w_i (f_i^* - f_{ij}) / (f_i^* - f_i^-) \right], \tag{14}$$

where w_i are the weights of criteria, expressing their relative importance.

Step 4-4: Obtain the performance index. Compute the values $Q_j, j = 1, 2, .., m$ by the relation.

$$Q_j = v(S_j - S^*) / (S^- - S^*) + (1 - v)(R_j - R^*) / (R^- - R^*) \tag{15}$$

where v is introduced as weight of the strategy of "the majority of criteria" (or "the maximum group utility"), here $v = 0.5$.

Step 4-5: Rank the alternatives for a compromise solution. Sorting by the values S_j, R_j and Q_j, in decreasing order. Propose as a compromise solution the alternative (a) which is ranked the best by the measure Q (minimum) . The best alternative, ranked by Q, is the one with the minimum value of Q. The main ranking result is the compromise ranking list of alternatives, and the compromise solution with the "advantage rate".

Outline of the Steps in the Analytical Methods

By summarizing the above analytical methods in Section 3.1 to 3.4, an outline of the steps can be listed below.

Modified Delphi Method

Step 1-1: DR and Big Data experts are invited to provide opinions on the criteria for evaluating and selecting the DR sites for information systems of academic Big Data.

Step 1-2: According to the definition of the Modified Delphi method introduced in Section 3.1, agreement by 2/3 (67%) of participants was taken as a threshold value for accepting a criterion.

DEMATEL

Step 2-1: DR and Big Data experts are asked to fill the questionnaires of pair-wise comparisons in terms of the influences and directions between the criteria. The initial direct relation/influence matrix ***A*** is derived then based on the inputs by the experts.

Step 2-2: The normalized matrix ***N*** is derived based on the initial direct relation/influence matrix ***A*** by Eqs. (1) and (2).

Step 2-3: The total relationship matrix ***T*** will be derived based on the normalized matrix ***N*** by Eq. (3).

Step 2-4: The row sums and column sums of the total relationship matrix will be calculated. Thus, the cause and effect relationships between criteria can be established. To simplify the decision problem, the criteria that receive influences only from other criteria will be discarded.

Step 2-5: The structure of the DR site evaluating and selecting problem will be established based on the total relationship matrix. Appropriate threshold value will be defined for simplifying the relationships between criteria. The DEMATEL method constructs the interrelations between factors/criteria to build a NRM.

DNP

Step 3-1: This study uses influence levels values of each element over others from NRM as the basis of the normalization supermatrix for determining ANP weights to obtain the relative importance.

Step 3-2: The unweighted supermatrix will be derived by Eq. (10).

Step 3-3: The weighted supermatrix will be raised to limiting powers to obtain the global priority vector (or called weights) by Eq. (11).

VIKOR

Step 4-1: Determine the positive-ideal level. The positive-ideal level can be determined using Eq. (12)

Step 4-2:Normalize the original rating matrix using Eq. (13).

Step 4-3:Compute the values S_j and R_j using Eq. (14).

Step 4-4:Obtain the performance index.

Step 4-5: Rank the alternatives for a compromise solution.

EVALUATION AND SELECTION OF A DISASTER RECOVERY SITE FOR BIG DATA APPLICATIONS

In this section, an empirical study on DR site evaluation and selection for Big Data applications is used to demonstrate the feasibility of the proposed method. First, the background and nature of the empirical study case in Asia is introduced. The experts were selected from the study case for evaluating and selecting a DR site. Then, the decision evaluation is structured based on the dimensions and factors derived by using the DEMATEL method. Then the key determinants for DR sites is derived using the DNP method. Finally, this research uses the VIKOR method with the DNP influence weights to select DR sites and provide improvement suggestions. These detailed assessment processes are demonstrated in the following sub-sections.

Background and Problem Description

Founded in 1991, Center X is the top and national data center in an Asian economy. Center X builds complete cloud computing, storage, and networking facilities for domestic and international academic institutes in three locations. Since its establishment, Center X has provided DR services for large-scale research and government institutes, which are characterized by high availability concerns, large-scale data backup and processing, as well as cutting-edge IT technology adoption. From grid computing to cloud computing, Center X has developed technologies that integrate high-bandwidth networks, high-performance computing (HPC), and highly efficient storage facilities that serve scientific research and applications in various fields. In recent years, computing and monitoring data gathered by Center X from the fields of earth science, biomedicine, and disaster prevention have exceeded 200TB data; the scale of data is still growing. To effectively use such Big Data, Center X developed its own Big Data processing and analysis technologies and utilized existing Big Data processing technologies like Hadoop. Center X also integrated its specialized knowledge within each field to maximize the value of Big Data to the society and environment.

There are three DR sites, site A (headquarters), site B (branch office), and site C (branch office), belonging to Center X. All the three sites provide remote backup mechanisms and reliable DR services. Center X has adopted the ISO 27001 and BS 7799 standards for security management to provide reliable DR services in the three sites. Center X has high-efficiency and high-capacity storage equipment at its three sites. The sites are linked by a dual optical fiber backbone with high bandwidth, which is supported by governments funding. The storage area network structure is used along with a dual backbone network characterized by uninterrupted data transmissions, which not only significantly increases the data backup speed, but also achieves rapid data recovery in case of losses or failures.

To verify the proposed analytic framework, this research chooses the three sites in Center X as the empirical study case. The experts were selected from the participants in 3 Big Data programs belonging to Center X, which include a data backup program for large-scale satellite images and aerial photograph as well as a system recovery program for the management of information systems to sustain daily operations. The experts selected included five IT managers who are responsible for DR operations in Center X and 5 project managers from the users of DR sites belonging to Center X. All the experts have more than five years of work experiences in the related fields of business continuity management and DR plan definition.

Network Relationship Map Construction using the DEMATEL

At first, the 5 dimensions and 22 factors being derived by the authors (Yang et al., 2015) were introduced. The dimensions and factors have been confirmed by the 10 experts (i.e. the 5 IT managers and 5 project managers being mentioned in Section 4.1) as suitable for evaluating and selecting DR sites by using the modified Delphi method in Section 3.1. All of the dimensions and factors were confirmed as appropriate. To derive the inter-influence relationships between the factors, opinions provided by experts were surveyed. The DR experts were asked to determine the influence relationships among the dimensions and factors. Then, using Eq. (1)-(4), the total relation matrix of dimension ($T_{dimensions}$) and factors ($T_{factors}$) were derived, as shown in **Table 2** and **Table 3**. Using Eq. (5) (6), the NRM was constructed by the r_i and c_i being derived from the total relation matrix, as shown in **Table 4**. The causal diagram consisting of the influence relationships between the dimensions and those influence relationships between each criterion belonging to each dimension are demonstrated in **Figure 2**. By referring to Yang et al. (2015), this research adopted the statistical method and set the threshold value as $\mu + \left(\frac{1}{2}\right)\delta$ of the total influence relationships.

In the NRM, the casual relationship incorporates five dimensions from **Figure 2**. IT managers may use the influence relationships to create the improvement strategy for DR sites. For example, IT System Availability (B) and DR Objectives (C), with the highest influence of $r_i + c_i$ as the prominence, represents both dimensions as core items for the DR site selection and evaluation problem. These two dimensions play central roles and need improvement as a top priory. However, IT System Availability (B) and DR Objectives (C) have negative scores in $r_i - c_i$, which represents that these two dimensions are the receivers affected by others. **Figure 2** demonstrates that DR Objectives (C) is affected by IT System Availability (B) and Disaster Readiness exercises (D). IT System Availability (B) is affected by all of the dimensions. This implies that when an organization enhancing its IT Facility Availability (B) and Disaster Readiness exercises (D) according to operating requirements and new technology adoption. It can positively affect the improvement of the DR Objectives (C) ability. Conversely, the dimensions with high relations and low prominence are Location and Infrastructure (A) and Operation management (E) due to the prominence being lower than the mean for other DR site consideration factors; the causal relationship has a small impact and improvement can be given a lower priority.

Using DNP to calculate the influence weights for each criterion

This study combines the DEMATEL technique with the ANP method to obtain the influential weights of the factors. Initially, the influence of the relationships among the factors was compared based on the NRM. An unweighted supermatrix W_c can be obtained by transposing the normalized matrix T^{nor}, as shown in **Table 5** to **Table 8**. Finally, the influential weights of DNP can be obtained by limiting the weighted super-matrix W_c^* until it reaches a steady state, as shown in **Table 10** to **Table 11**.

Table 2. Direct relation/influence matrix $A_{dimensions}$ of dimensions

	0.000	2.500	2.200	2.400	1.700
	2.000	0.000	3.100	2.600	2.100
$A_{\text{dimensions}}$ =	2.000	3.000	0.000	2.500	2.400
	2.400	2.600	2.500	0.000	2.400
	1.600	2.300	2.600	2.200	0.000

Table 3. Direct relation/influence matrix $A_{factors}$ of factors

	0.000	2.100	2.700	2.300	2.400	2.100																
	1.100	0.000	2.400	1.800	2.600	2.100																
	2.400	1.800	0.000	2.400	1.500	1.500																
	1.700	1.700	2.900	0.000	1.400	1.300																
	2.000	2.200	1.200	1.300	0.000	2.600																
	2.300	2.000	1.600	1.000	3.100	0.000																
							0.000	2.500	2.900	3.000	2.300											
							1.800	0.000	2.500	2.100	1.800											
							2.300	2.400	0.000	2.700	2.000											
							2.800	2.200	3.100	0.000	2.900											
$A_{factors}$ =							1.800	1.700	2.000	3.000	0.000											
												0.000	3.000	2.300								
												3.100	0.000	2.300								
												2.500	2.900	0.000								
															0.000	2.300	2.500					
															2.300	0.000	2.500					
															2.200	2.200	0.000					
																		0.000	2.400	2.800	2.700	2.100
																		2.100	0.000	2.500	2.600	1.900
																		2.600	2.400	0.000	2.600	2.300
																		2.800	2.800	3.000	0.000	2.900
																		2.000	1.800	2.300	2.900	0.000

The influential weights corresponding to each dimension and criterion can thus be derived based on the weighted super matrices shown in **Table 11**. According to **Table 11**, the dimensions can be prioritized as (1) IT Facility Availability (B), (2) DR Objectives (C), (3) Disaster Readiness Exercises (D), (4) Operation management (E), and (5) Location and Infrastructure (A). Furthermore, as for the most important factors, the top-ranking factors include (1) Recovery Time Objective (c_2), (2) Recovery Point Objective (c_1), (3) Emergency Operations Center (d_3), (4) Education and Training (d_1), and (5) DR Work Area (d_2). The result demonstrates that Recovery Time Objective (c_2) and Recovery Point Objective (c_1) have the most influence weight and need to be considered in DR site evaluation and selection. All these influence weights are used in selecting the best alternatives in MCDM problems with VIKOR.

Table 4. Normalized direct relation/influence matrix $N_{dimensions}$ of dimensions

	2.541	2.106	2.090	1.992	1.783
	1.846	3.093	2.326	2.174	1.965
$N_{diemensions}$ =	1.855	2.329	3.109	2.180	1.994
	1.877	2.298	2.296	2.980	1.988
	1.660	2.076	2.099	1.963	2.628

Table 5. Normalized direct relation/influence matrix $N_{factors}$ of factors

	1.948	1.123	1.229	1.049	1.237	1.109															
	0.918	1.848	1.074	0.901	1.116	0.991															
	0.981	0.963	1.895	0.931	1.025	0.931															
	0.891	0.907	1.043	1.716	0.962	0.868															
	0.937	0.972	0.961	0.834	1.901	0.989															
	1.007	1.012	1.037	0.863	1.170	1.858															
							1.840	1.017	1.176	1.199	1.025										
							0.803	1.669	0.954	0.944	0.819										
							0.919	0.926	1.876	1.080	0.919										
							1.050	1.016	1.206	2.015	1.077										
							0.833	0.825	0.958	1.032	1.716										
$N_{factors}$ =												1.167	0.378	0.306							
												0.381	1.176	0.309							
												0.347	0.375	1.143							
															1.112	0.277	0.299				
															0.277	1.112	0.299				
															0.263	0.263	1.114				
																		1.000	0.000	0.000	0.000	0.000
																		0.545	1.295	0.498	0.512	0.439
																		0.592	0.475	1.331	0.523	0.473
																		0.661	0.543	0.587	1.391	0.553
																		0.540	0.431	0.488	0.531	1.300

Table 6. Total relation matrix $T_{dimensions}$ of dimensions

	1.541	2.106	2.090	1.992	1.783
	1.846	2.093	2.326	2.174	1.965
$T_{\dim ensions}$ =	1.855	2.329	2.109	2.180	1.994
	1.877	2.298	2.296	1.980	1.988
	1.660	2.076	2.099	1.963	1.628

Table 7. Total relation matrix $T_{factors}$ of factors

	0.948	1.123	1.229	1.049	1.237	1.109	1.872	1.875	2.177	2.219	1.919	0.623	0.647	0.528	0.433	0.433	0.473	2.059	2.039	2.247	2.273	2.003
	0.918	0.848	1.074	0.901	1.116	0.991	1.872	1.875	2.177	2.219	1.919	0.623	0.647	0.528	0.433	0.433	0.473	2.059	2.039	2.	2.273	2.003
	0.981	0.963	0.895	0.931	1.025	0.931	1.872	1.875	2.177	2.219	1.919	0.623	0.647	0.528	0.433	0.433	0.473	2.059	2.039	2.	2.273	2.003
	0.891	0.907	1.043	0.716	0.962	0.868	1.872	1.875	2.177	2.219	1.919	0.623	0.647	0.528	0.433	0.433	0.473	2.059	2.039	2.	2.273	2.003
	0.937	0.972	0.961	0.834	0.901	0.989	1.872	1.875	2.177	2.219	1.919	0.623	0.647	0.528	0.433	0.433	0.473	2.059	2.039	2.	2.273	2.003
	1.007	1.012	1.037	0.863	1.170	0.858	1.872	1.875	2.177	2.219	1.919	0.623	0.647	0.528	0.433	0.433	0.473	2.059	2.039	2.	2.273	2.003
	1.749	1.792	1.919	1.629	1.973	1.768	0.840	1.017	1.176	1.199	1.025	0.694	0.721	0.588	0.473	0.473	0.516	2.269	2.248	2.	2.505	2.208
	1.749	1.792	1.919	1.629	1.973	1.768	0.803	0.669	0.954	0.944	0.819	0.694	0.721	0.588	0.473	0.473	0.516	2.269	2.248	2.	2.505	2.208
	1.749	1.792	1.919	1.629	1.973	1.768	0.919	0.926	0.876	1.080	0.919	0.694	0.721	0.588	0.473	0.473	0.516	2.269	2.248	2.	2.505	2.208
	1.749	1.792	1.919	1.629	1.973	1.768	1.050	1.016	1.206	1.015	1.077	0.694	0.721	0.588	0.473	0.473	0.516	2.269	2.248	2.	2.505	2.208
$T_{factors}$ =	1.749	1.792	1.919	1.629	1.973	1.768	0.833	0.825	0.958	1.032	0.716	0.694	0.721	0.588	0.473	0.473	0.516	2.269	2.248	2.	2.505	2.208
	1.757	1.801	1.929	1.637	1.983	1.777	2.070	2.074	2.408	2.454	2.122	0.167	0.378	0.306	0.474	0.474	0.517	2.303	2.281	2.	2.542	2.241
	1.757	1.801	1.929	1.637	1.983	1.777	2.070	2.074	2.408	2.454	2.122	0.381	0.176	0.309	0.474	0.474	0.517	2.303	2.281	2.	2.542	2.241
	1.757	1.801	1.929	1.637	1.983	1.777	2.070	2.074	2.408	2.454	2.122	0.347	0.375	0.143	0.474	0.474	0.517	2.303	2.281	2.	2.542	2.241
	1.778	1.822	1.951	1.656	2.006	1.798	2.043	2.047	2.376	2.421	2.094	0.685	0.711	0.580	0.112	0.277	0.299	2.296	2.275	2.	2.535	2.234
	1.778	1.822	1.951	1.656	2.006	1.798	2.043	2.047	2.376	2.421	2.094	0.685	0.711	0.580	0.277	0.112	0.299	2.296	2.275	2.	2.535	2.234
	1.778	1.822	1.951	1.656	2.006	1.798	2.043	2.047	2.376	2.421	2.094	0.685	0.711	0.580	0.263	0.263	0.114	2.296	2.275	2.	2.535	2.234
	1.572	1.611	1.726	1.465	1.774	1.590	1.846	1.849	2.147	2.188	1.892	0.626	0.650	0.530	0.456	0.456	0.498	1.030	1.191	1.	1.323	1.155
	1.572	1.611	1.726	1.465	1.774	1.590	1.846	1.849	2.147	2.188	1.892	0.626	0.650	0.530	0.456	0.456	0.498	1.107	0.944	1.	1.233	1.068
	1.572	1.611	1.726	1.465	1.774	1.590	1.846	1.849	2.147	2.188	1.892	0.626	0.650	0.530	0.456	0.456	0.498	1.202	1.180	1.	1.307	1.156
	1.572	1.611	1.726	1.465	1.774	1.590	1.846	1.849	2.147	2.188	1.892	0.626	0.650	0.530	0.456	0.456	0.498	1.342	1.330	1.	1.266	1.316
	1.572	1.611	1.726	1.465	1.774	1.590	1.846	1.849	2.147	2.188	1.892	0.626	0.650	0.530	0.456	0.456	0.498	1.095	1.074	1.	1.245	0.923

Table 8. $r_i + c_i$ and $r_i - c_i$ values calculated from the direct/indirect matrix ***T***.

Dimensions	A	B	C	D	E
$r_i + c_i$	18.291	21.306	21.386	20.728	18.783
$r_i - c_i$	0.732	-0.498	-0.453	0.151	0.068

Factors	***a₁***	***a₂***	***a₃***	***a₄***	***a₅***	***a₆***	***b₁***	***b₂***	***b₃***	***b₄***	***b₅***
$r_i + c_i$	12.377	11.672	11.965	10.681	12.005	11.695	9.702	8.641	9.890	10.632	8.920
$r_i - c_i$	1.011	0.024	-0.511	0.094	-0.817	0.199	0.813	-0.265	-0.450	0.095	-0.193

Factors	***c₁***	***c₂***	***c₃***	***d₁***	***d₂***	***d₃***	***e₁***	***e₂***	***e₃***	***e₄***	***e₅***
$r_i + c_i$	1.747	1.794	1.624	1.341	1.341	1.353	11.791	11.288	12.259	13.087	11.155
$r_i - c_i$	-0.042	-0.065	0.107	0.036	0.036	-0.072	0.242	-0.151	-0.347	0.338	-0.082

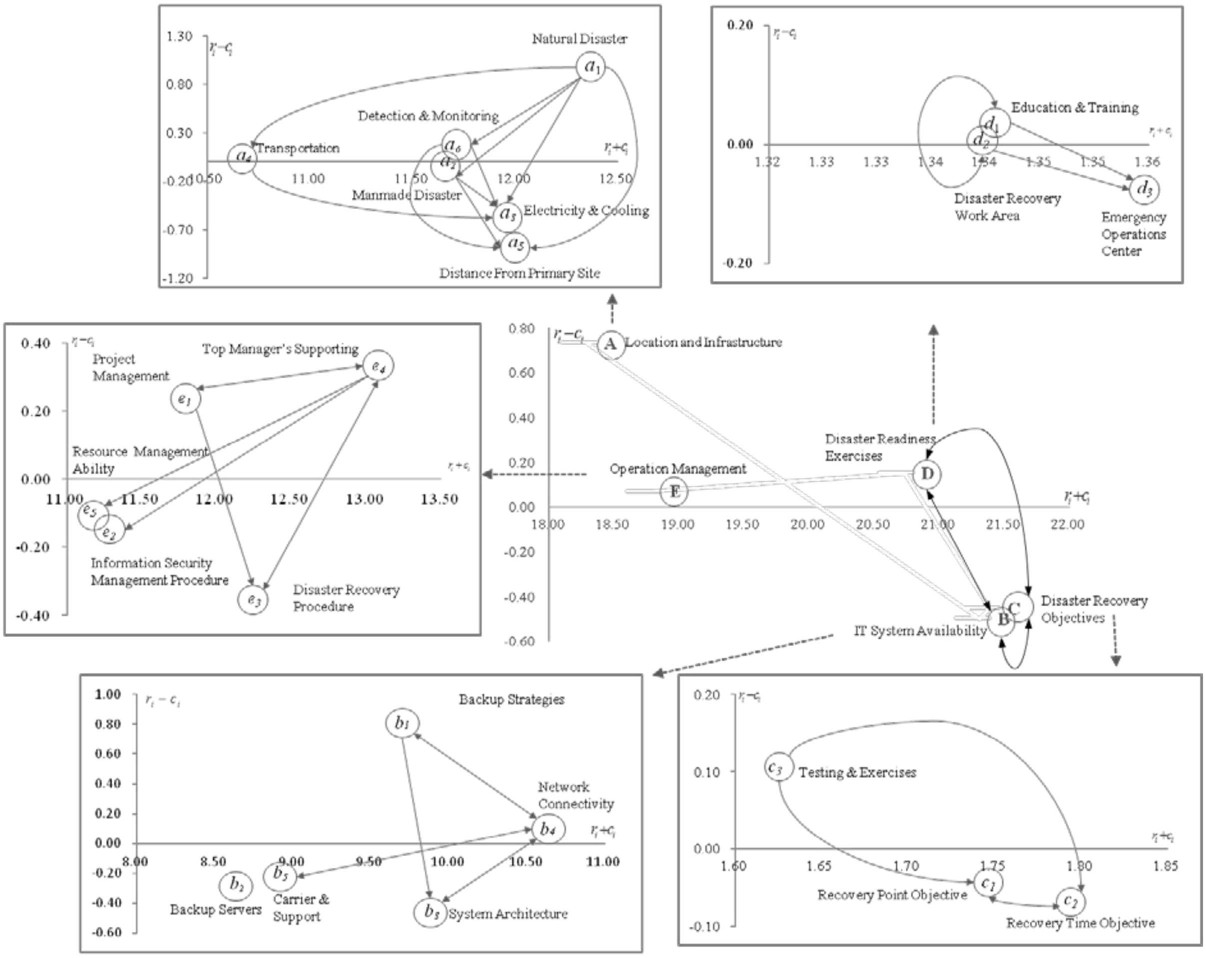

Figure 2. Causal diagram of total relationship. Note: The threshold is set at $\mu + (1/2)\delta$ of the total relationships

Table 9. Unweighted Supermatrix W_c of factors.

$W_{c\_factors}$=

0.023	0.025	0.028	0.027	0.027	0.027	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.02	0.028	0.028	0.028	0.028	0.028
0.027	0.023	0.027	0.027	0.028	0.028	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.030	0.030	0.03	0.029	0.029	0.029	0.029	0.029
0.030	0.030	0.025	0.031	0.028	0.028	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.032	0.032	0.03	0.031	0.031	0.031	0.031	0.031
0.025	0.025	0.026	0.022	0.024	0.023	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.02	0.026	0.026	0.026	0.026	0.026
0.030	0.031	0.029	0.029	0.026	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.033	0.033	0.03	0.032	0.032	0.032	0.032	0.032
0.027	0.027	0.026	0.026	0.029	0.023	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.02	0.029	0.029	0.029	0.029	0.029
0.041	0.041	0.041	0.041	0.041	0.041	0.032	0.039	0.039	0.039	0.038	0.041	0.041	0.041	0.041	0.041	0.04	0.041	0.041	0.041	0.041	0.041
0.041	0.041	0.041	0.041	0.041	0.041	0.039	0.032	0.039	0.038	0.038	0.046	0.046	0.046	0.045	0.045	0.04	0.040	0.040	0.040	0.040	0.040
0.048	0.048	0.048	0.048	0.048	0.048	0.045	0.046	0.037	0.045	0.044	0.048	0.048	0.048	0.048	0.048	0.04	0.048	0.048	0.048	0.048	0.048
0.049	0.049	0.049	0.049	0.049	0.049	0.046	0.045	0.046	0.038	0.048	0.049	0.049	0.049	0.049	0.049	0.04	0.049	0.049	0.049	0.049	0.049
0.042	0.042	0.042	0.042	0.042	0.042	0.039	0.039	0.039	0.040	0.033	0.042	0.042	0.042	0.042	0.042	0.04	0.042	0.042	0.042	0.042	0.042
0.076	0.076	0.076	0.076	0.076	0.076	0.077	0.077	0.077	0.077	0.077	0.040	0.089	0.081	0.076	0.076	0.07	0.077	0.077	0.077	0.077	0.077
0.079	0.079	0.079	0.079	0.079	0.079	0.080	0.080	0.080	0.080	0.080	0.089	0.041	0.087	0.079	0.079	0.07	0.080	0.080	0.080	0.080	0.080
0.065	0.065	0.065	0.065	0.065	0.065	0.066	0.066	0.066	0.066	0.066	0.072	0.072	0.033	0.065	0.065	0.06	0.065	0.065	0.065	0.065	0.065
0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.067	0.067	0.067	0.031	0.076	0.07	0.067	0.067	0.067	0.067	0.067
0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.067	0.067	0.067	0.076	0.031	0.07	0.067	0.067	0.067	0.067	0.067
0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.082	0.082	0.03	0.074	0.074	0.074	0.074	0.074
0.036	0.036	0.036	0.036	0.036	0.036	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.03	0.030	0.034	0.035	0.035	0.034
0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.037	0.037	0.037	0.037	0.037	0.03	0.034	0.029	0.034	0.034	0.033
0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.04	0.038	0.038	0.032	0.038	0.037
0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.041	0.041	0.041	0.041	0.041	0.04	0.038	0.038	0.038	0.033	0.039
0.035	0.035	0.035	0.035	0.035	0.035	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.03	0.033	0.033	0.034	0.034	0.029

Table 10. Limited Supermatrix W_c^* of Criteria

$W*_{c\_factors}=$

0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.02	0.028	0.028	0.028	0.028	0.028
0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.02	0.029	0.029	0.029	0.029	0.029
0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.03	0.031	0.031	0.031	0.031	0.031
0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.02	0.026	0.026	0.026	0.026	0.026
0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.03	0.032	0.032	0.032	0.032	0.032
0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.02	0.029	0.029	0.029	0.029	0.029
0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.041	0.041	0.041	0.041	0.041	0.04	0.040	0.040	0.040	0.040	0.040
0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.04	0.042	0.042	0.042	0.042	0.042
0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.04	0.047	0.047	0.047	0.047	0.047
0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.04	0.048	0.048	0.048	0.048	0.048
0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.04	0.041	0.041	0.041	0.041	0.041
0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.077	0.075	0.075	0.075	0.075	0.07	0.075	0.075	0.075	0.075	0.075
0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.080	0.078	0.078	0.078	0.07	0.078	0.078	0.078	0.078	0.078
0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.065	0.064	0.064	0.06	0.064	0.064	0.064	0.064	0.064
0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.067	0.067	0.067	0.068	0.066	0.06	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	0.067	0.067	0.067	0.066	0.068	0.06	0.066	0.066	0.066	0.066	0.066
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	0.07	0.072	0.072	0.072	0.072	0.072
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	0.03	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	0.036	0.036	0.036	0.036	0.03	0.036	0.036	0.036	0.036	0.036
0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.040	0.040	0.040	0.040	0.040	0.04	0.039	0.039	0.039	0.039	0.039
0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.04	0.040	0.040	0.040	0.040	0.040
0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.03	0.035	0.035	0.035	0.035	0.035

Table 11. Performance and gaps with respect to aspired level of alternative disaster recovery site

Aspects and factors	Weights	Performance			Gaps to the aspiration (VIKOR)		
		Site A	Site B	Site C	Site A	Site B	Site C
Location and Infrastructure (A)	0.175 (5)						
Natural Disaster (a_1)	0.028	3.050	3.350	3.350	0.028	0.019	0.019
Manmade Disaster (a_2)	0.029	3.050	3.450	3.500	0.029	0.017	0.015
Distance from Primary Site (a_3)	0.031	3.100	3.200	3.100	0.031	0.027	0.031
Transportation (a_4)	0.026	3.900	2.700	2.400	0.002	0.021	0.026
Electricity and Cooling (a_5)	0.032	3.050	3.250	3.350	0.032	0.025	0.022
Detection and Monitoring (a_6)	0.029	3.550	3.250	3.550	0.017	0.028	0.017
IT System Availability (B)	0.219 (1)						
Backup Strategies (b_1)	0.040	3.500	3.300	3.200	0.024	0.034	0.039
Backup Servers (b_2)	0.042	3.300	3.350	3.200	0.034	0.031	0.039
Backup System	0.047	3.250	3.350	3.000	0.034	0.029	0.045
Network Connectivity (b_4)	0.048	3.600	3.450	3.600	0.033	0.046	0.033
Carrier and Support (b_5)	0.041	3.400	3.150	3.350	0.028	0.040	0.030
DR Objectives (C)	0.217 (2)						
Recovery Point Objective (c_1)	0.075	3.050	3.100	2.850	0.064	0.061	0.078
Recovery Time Objective (c_2)	0.078	3.050	3.000	2.800	0.063	0.067	0.080
Testing and Exercises(c_3)	0.064	3.250	3.050	3.050	0.053	0.068	0.068
Disaster Readiness exercises (D)	0.205 (3)						
Education and Training(d_1)	0.066	3.100	2.750	2.600	0.040	0.056	0.062
DR Work Area(d_2)	0.066	2.700	2.550	2.500	0.054	0.060	0.062
Emergency Operations Center	0.072	2.850	2.900	2.650	0.056	0.054	0.066
Operation management (E)	0.186 (4)						
Project Management (e_1)	0.036	3.350	3.000	3.250	0.027	0.037	0.039
Information Security	0.036	3.150	2.950	2.900	0.025	0.039	0.029
DR Procedure (e_3)	0.039	3.450	3.050	3.100	0.033	0.041	0.043
Top Manager's Supporting (e_4)	0.040	3.450	2.850	2.750	0.025	0.043	0.041
Resources Management (e_5)	0.035	3.350	3.000	3.250	0.017	0.035	0.038
Total	1.000	71.450 (1)	68.050 (2)	67.050 (3)	N.A.	N.A.	N.A.
Total gap S_i	N.A.	N.A.	N.A.	N.A.	0.717	0.832	0.889
Gap R_i	N.A.	N.A.	N.A.	N.A.	0.064	0.068	0.080
Qj (v=0.5)	N.A.	N.A.	N.A.	N.A.	0.000 (1)	0.447 (2)	1.000 (3)

Remarks: 1. N.A. means not applicable.
2. The numbers in the parenthesis stand for the ranking of the number.

Using the VIKOR method to evaluate DR site performance and gaps

This study applies the VIKOR method to derive the compromise rankings after calculating the influence weights for the factors using DNP. Based on the influence weights derived by the DNP introduced in Section 3, we combine the DNP with the VIKOR method to obtain the total performance scores for site A, B, and C, which were 71.450, 68.050, and 67.050, respectively (**Table 11**). We use Eqs. (12)-(15) to derive the gaps between the current status and the aspired level of each criterion belonging to each alternative. The gaps to the aspired level are site A (0.000) $\prec$ site B (0.447) $\prec$ site C (1.000). The results are detailed in **Table 11**.

The results indicate the improvement priority sequence necessary for the overall factors to reach the desired level. The high-performance value reveals that the factors have been regarded as more important and received sufficient support from the organizations. Consequently, high performance value has a lower gap value to the optimal scale and becomes the easy start for decision-makers to make improvements if their goal is set to reach the aspired level. For example, in site A, Recovery Point Objective (c_1), with a higher influence value of 0.075 and larger gap value of 0.64, indicates that improvements are prioritized to reach the aspiration level. This is followed by Recovery Time Objective (c_2) and Emergency Operation Center (d_3). In site B, Testing and Exercises (c_3) and Recovery Time Objective (c_2) have a larger gap value of 0.68 and 0.067, indicating improvements are prioritized to reach the aspired level. In site C, Recovery Time Objective (c_2), Recovery Point Objective (c_1), and Testing and Exercises (c_3) have higher gap values than others and are apparently the first factors on to be improved to reach the desired level.

DISCUSSION

To resolve the complex IT DR site evaluation, selection, and enhancement problem, this study has established an analytic framework to reduce the gaps between the current status and the aspired level versus each criterion. In Section 4, an empirical study based on a real Big Data DR application is used to illustrate the feasibility of the proposed framework. The key determinants for IT DR site selection have been derived by using the proposed hybrid MCDM framework for the best DR site selection. In this section, the analytic results are discussed from both dimensions of managerial implications in DR site selection, evaluation, and improvement strategy for the individual site.

Managerial Implications

This study derived the relative influence weights among dimensions and factors (refer **Table 11**). It is extremely important to deliver such relative importance information to IT managers because the parameters can contribute to clarify what factors are more important for the Big Data DR site. Then, this research combined the DEMATEL results (refer **Figure 2**) to understand what dimensions and factors influence the more important factors being identified by the DNP results. Such an understanding enables managers to define DR site-improving strategies for resolving the problems from the root causes. In the following sub-sections, the causal relationships are discussed.

Prioritization of the Dimensions and the Influence Relationships between Dimensions

According to the DNP results (**Table 11**), the IT System Availability (B) and DR Objectives (C) were ranked as the most important dimensions with an influence weight of 0.219 and 0.217, respectively. That is, the IT System Availability (B) dimension is more important than the DR Objectives (C) dimension. This result is not consistent with the previous studies on DR site selection (Wang & Pin, 2008; Yang et al., 2015).

In general, the DR Objectives (C) dimension is more important than IT System Availability (B) because the DR objectives (C) dimension is the key performance indicator for business continuity plans. Once the DR objectives have been defined, the IT systems are designed to meet the requirements of the DR objectives (Broder & Tucker, 2011). However, when the scale, scope, and velocity of Big Data exceed the capabilities of traditional IT systems, IT systems become the major uncertainty and constraint in choosing a DR site or setting DR objectives. To cope with such situations, organizations are devoted to the design and architecture of future hardware and software systems for the Big Data application (Kambatla, Kollias, Kumar, & Grama, 2014). For example, more organizations are considering the use of flexible and scalable cloud computing architecture to fulfill the needs from both their data storage requirements and the heavy server processing required to analyze large volumes of data economically (Villars, Olofson, & Eastwood, 2011). Therefore, in the era of Big Data, the IT System Availability (B) plays a more dominant role than the DR objectives (C).

DR Objectives (C) is the second most important dimension. This result is tied closely to the overall business continuity management system process. For example, international Standard BS25999, the BSI's standard in the field of Business Continuity Management, requires organizations to define an effective business continuity management system that emphasizes the importance of business continuity needs and establishes the objectives for business continuity (BS25999, 2007). Defining DR Objectives is the key step in developing business continuity plan for mapping business processes and associated resources and establishing appropriate monitoring and control mechanisms. Since the establishment of a viable DR system for business continuity is important to mitigate risks (Han et al., 2012), once the risks and their impacts have been defined, the DR technologies can be selected to meet the data protection requirements of the organization (Bertrand, 2005). Therefore, DR objectives (C) ties closely to IT System Availability (B) and Disaster Readiness exercises (D) because of the resumption of full operations after a period of recovery time that requires data, platforms, applications, and people. For example, if the DR Objectives require a recovery time close to zero, the IT systems should automatically switch to a redundant system to minimize IT service interruption time. Implementations of such mechanisms need investments in a lot of equipment rapidly. Thus, the DR site becomes very expensive. However, if an organization has more tolerance for service downtime and data losses, the recovery time for the IT system and data can be longer; thus, the DR site could use traditional tape backup approach with lower costs (Wiboonrat, 2008). Our research also demonstrated the inter-influence relationships among DR Objectives (C), IT System Availability (B), and Disaster Readiness Exercises (D), as shown in **Figure 2**.

The Disaster Readiness Exercises (D) and Operation Management (E) were ranked as the third and fourth most important dimensions, with an influence weight of 0.205 and 0.186, respectively. Disaster Readiness Exercises (D) is part of emergency management. The purpose of Disaster Readiness Exercises (D) is to support the emergency response operations across organizational, jurisdictional, and geographic bodies. Many organizations have established the emergency response standard operating procedures (SOPs). For example, the National Incident Management System (NIMS) was established by a Homeland Security Presidential Directive of the U.S. government. This system prescribes institutional response guidelines in establishing rule structures and developing a normative environment with defined tasks regarding what should be done during a response. However, once a complex and uncertain disaster occurs, responders must make rapid coordination decisions, which constrain their capabilities to analyze coordination problems and explore the solution domain besides the SOPs. Conversely, Operation Management (E) belongs to the DR planning process, which is often outsourced as a separate project. Organizations usually use project management methodologies to manage the process and resources. A plan-based approach relies heavily on SOPs to address coordinated planning and training exercises involving stakeholders (Chen, Sharman, Rao, & Upadhyaya, 2008). Therefore, the Disaster Readiness Exercises (D) dimension plays a more dominant role than the Operation Management (E) dimension.

Location and Infrastructure (A) was ranked as the least important dimension for the DR site evaluation and selection problem. Apparently, this result is also not consistent with previous studies of data center selection (Balcik & Beamon, 2008; Görmez et al., 2011). Most of the earlier works focused on site location and were devoted to developing various programming algorithms to optimize the DR site capacity and locate facilities on nodes to minimize costs (Hassin et al., 2010; Miyagawa, 2012; Pirkul & Schilling, 1989). Our research results suggested a different viewpoint from the dimension of Big Data DR sites. According to Altay and Green (2006), 109 journal articles were published in the operation research field of emergency management from 1980 through 2006. They observed that only 11.9% of articles focused on natural disasters. The rest covered manmade emergencies (47.6%) and general approaches to emergency management (40.5%) designed to apply to all disaster situations. Unlike natural disasters, manmade disasters such as hacker attacks via the internet cannot be mitigated by geographic isolation to prevent occupation of high hazard areas. Due to the nature of large volume and rapid changes, Big Data is impacted more by manmade disasters than natural disasters. In our results of **Table 11**, the man-made disasters (a_2) was ranked higher than the natural disasters (a_1). Since man-made disasters are the mainstream disaster type for Big Data, the location and infrastructure become less important. That is why Location and Infrastructure (A) was ranked as the least important dimension for the DR site evaluation and selection problem in the era of Big Data.

Factors in the IT System Availability (B) Dimension

Based on the analytic results being demonstrated in **Table 11**, Network Connectivity (b_4) is the most important criterion in the IT System Availability (B) dimension for a DR site selection problem for a Big Data application. This is because network connectivity has become a major component of the modern data center. This result is consistent with previous studies discussing the infrastructure of Big Data, which have argued that the network must be a key investment (Gantz & Reinsel, 2012). The first reason for the analytic result is the requirement for sufficient network bandwidth for Big Data applications. Due to the Big Data characteristics of volume and velocity, the terabytes of data set consume large amounts of network bandwidth for uploading such Big Data to the DR site. Therefore, insufficient network capacity can become the bottleneck in data storage and recovery systems in DR sites (Zeng et al.,2015; Katal, Wazid, & Goudar, 2013; Zhang, Yan, Xu, & Su, 2014). The second reason is the considerations of fault tolerance, security, and access control of networks. Such considerations are important for some critical businesses. Some DR sites need dual networks to fulfill the DR objectives that must be met. The cost for establishing the network infrastructure is doubled. The third reason is that Big Data are generated, collected, and distributed from multiple sources. DR sites need more network connections from various remote data centers. The existing routing strategy between the primary site and DR sites fails to exploit the link diversity of data center networks. The more links are used; the higher bandwidth is required. Apparently, network bandwidth must be expanded as soon as the DR site begins to operate to generate speed in backup. In order to maximize the utilization rate of the network bandwidth, researchers must develop advanced network technologies with the most cost-effective solution(s) (Greenberg et al., 2009). Typical examples include the adoption of virtualization technology on networks as the software defined networks (SDNs), the putting of computation and storage in the same Hadoop node platform for Big Data applications to reduce the interconnection transactions, etc.

For the influence relationships in the IT System Availability (B) dimension, the Network and Connectivity (b_4) criterion plays a central role in the IT System Availability (B) dimension, with strong influence on Backup Strategies (b_1), Backup System Architecture (b_3), and Carrier and Support(b_5) (refer **Figure 2**). Apparently, the network capacity (b_4) limits the higher level of backup strategies (b_1) and architecture (b_3), e.g. the full backup and the hot site for Big Data applications. This paper has demonstrated that networks (b_4) are more important than servers (b_2) and storage concerns (b_1). Thus, IT managers being in charge of DR site selections for Big Data applications should treat the network bandwidth (b_4) as the key to a successful DR site selection for Big Data applications.

Factors in the DR Objectives (C) Dimension

Based on the analytic results demonstrated in **Table 11**, the Recovery Time Objective (RTO) (c_2) and Recovery Point Objective (RPO) (c_1) are the most important factors in the DR Objectives (C) dimension. The results are consistent with the work by Claunch (2004): RTO and RPO are the major indexes of DR Objectives. Both indexes refer to the assurance of business continuity within a maximum tolerable period of disruption. The RTO and RPO are defined by a business impact analysis process which verifies the mission-critical application base on regulation and customer issues. Therefore, organizations usually use the RPO and RTO as indexes to quantify their capability to respond for a crisis. However, such indexes depend on data protection and impact minimization capabilities when disruptions happen (Bertrand, 2005), the usage and importance of Big Data are difficult to define. An earlier work by the Yang et al. (2015) also asserted that RTO and RPO are important for DR. However, Big Data prompts IT managers to rethink about the appropriateness of applying the two indicators due to the cost issues.

Some researchers argued that Big Data are designed to extract values from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and analysis. The main applications of Big Data include analytics, data mining, and decision making. Big Data may not be a mission-critical or urgent business (Gantz & Reinsel, 2011). Therefore, the DR site for Big Data applications is less important. However, some researchers have argued that a DR site for Big Data applications is critical (Boyd & Crawford, 2012; Lyon, 2014). For example, one of the most important applications of Big Data analytics is health and human welfare. Such Big Data in the health and human welfare related fields are dominated by medical image data. Such image data is mainly consisted of personal genomics and high-throughput medical images. The data volume involves very large-scale integration; such Big Data is very hard to be backed up or recovered. Furthermore, the DR is necessary and critical in the health-related fields because of the needs to analyze diseases such as Ebola, MARS, H5N1, etc. For example, the U.S. Department of Defense and West African's Ministries of Health (MOH) used more than 10 terabytes of data set to make the short-term forecasts on the diffusion of Ebola. In such analysis of infectious disease, the DR is critical because such data is too valuable to be lost.

Therefore, the RPO and RTO must clearly define the risk exposures and loss expectancies of organizations. Because Big Data DR sometimes restores peta bytes of data sets, the cost of DR sites is increasing quickly. IT Researchers have discussed possible alternatives to minimize the cost of Big Data DR. Various novel information technologies, such as compression, reduplication, wide area network (WAN) accelerators, and faster storage technology are evolving rapidly to minimize the Big Data DR RTO and RPO. These information technologies are changing the critical database into Big Data DR assets which can be backed up easily (Lampa, Dahlö, Olason, Hagberg, & Spjuth, 2013).

Finally, based on the causal relationships being demonstrated in **Figure 2**, the RTO (c_2) is affected by Testing and Exercise (c_3) as well as the RPO (c_1). From the dimension of a data center, the RTO (c_2) is more complex and difficult to achieve than the RPO. Thus, data centers must define key RTOs: What systems are synchronized? When are they synchronized? Where are the data sources to be recovered? These are generic DR issues; however, these issues are becoming more complex in the Big Data era. As data centers promise their customers to meet the RTO, a wide variety of stake holders (e.g. executives, IT managers, store managers, register assistants, and software developers) must understand such causal relationships and join the testing and exercise.

Factors in the Disaster Readiness Exercises (D) Dimension

Based on the analytic results demonstrated in **Table 11**, Emergency Operations Center (EOC) (d_3) has the highest influence weight in the Disaster Readiness Exercises (D) dimension. This result is consistent with the results of previous studies on emergency response management. Emergency response management enables and supports emergency response operations across organizational, jurisdictional, and geographic bodies. The EOC deals with the strategic issues and works with a global picture, leveraging external resources to help on-site response. The actions of the EOC are based on a reflective and proactive posture and the EOC commanders typically operate with a large time window (Chen et al., 2008). Also, the DR planning process is contracted as a project based SOPs. The SOP is the pre-incident preparedness and leads to response inflexibility in the face of unexpected events. Variants in a disaster originating from hazard uncertainty, including incident, informational, and task flow, organizational structure, and environment, are hard to manage by SOP. To support fast response during complex incidents and uncertainties, responders must make rapid coordination decisions, which constrain their capabilities to analyze coordination problems and explore the solution domain. Response to disasters can be viewed as consisting of an onsite response coordinating entity and a remote management entity in an EOC (Chen et al., 2008). The coordination of EOC is demanding as it involves requirements typical of an emergency situation that includes high uncertainty and necessity for rapid decision making and response under temporal and resource constraints.

The coordination of emergency responses is challenging in DR sites because such a coordination work usually involves various departments (e.g. IT, customer service, etc. which can happen in an emergency; e.g. a disruption of the infrastructure support which is required for coordinating electricity, telecommunications, and transportation). This kind of coordination works is complicated by factors such as infrastructure interdependencies, multi-authority and massive personal involvement, conflicts of interest, as well as high demands for timely information. Therefore, the EOC is influenced by Education and Training (d_1) and DR work area (d_2), respectively (**Figure 2**). In addition, more organizations are considering the use of flexible and scalable cloud computation technologies for DR applications on multiple data centers being located at different geographic areas. Therefore, IT systems which can coordinate load distributions among different data centers are emerging. An appropriate EOC design and operation can complement the SOP to achieve reasonable quality of services in a DR readiness exercise.

Factors in the Operation Management Objectives (E) Dimension

From the results shown in **Table 11**, Top Manager's Support (e_4) is the most important criterion in the Operation Management Objective (E) dimension; the corresponding influence weight is 0.040. These results are consistent with past researches in information management. Naranjo-Gil (2009) analyzed the role of top management teams and management information systems (MIS). While implementing a DR system, the strategic goals of every department differ. Because multiple organizational goals shall be fulfilled, supports from the top management can coordinate the organization. Such coordination by top management can maximize organizational performance. Järveläinen (2013) found that top management support is a key factor for organizational alertness and preparedness; business continuity practices affect the perceived business impacts of information security management. Apparently, supports by top managements have a positive effect on organizational alertness and preparedness in continuity practices. A purely technical approach is insufficient because commitment to and awareness of continuity are required on every organizational level.

DR planning must yield a comprehensive plan that covers every critical dimension of the business. Therefore, the DR operation process is recognized as a separate project being managed by using project management methodologies or as one part of an integrated business continuity plan, which needs to coordinate all the key areas within the organization. According to the Snedaker (2013), executive supports for IT project is the top success factors. One reason is that DR projects need authorizations from top managers to re-prioritize the project. Also, top managers are more business-centric and finance-centric than technology-centric.

Factors in the Location and Infrastructure (A) Dimension

Based on the analytic results demonstrated in **Table 11**, Electricity and Cooling (a_5) and Distance from Primary Site (a_3) are the most important factors in the Location and Infrastructure (A) dimension. As mentioned above, for Big Data applications, Manmade Disasters (a_2) are more influential than Natural Disasters (a_1). The analytic result implies that geographic isolation to prevent high hazard occupation may not be ranked as a top priory in modelling a DR site decision making problem for Big Data applications. As what has been discussed in Section 5.1.1, the IT System Availability (B) dimension is more important that the DR Objectives (C) dimension. Our research results suggested a deeper viewpoint from the dimension of Big Data DR site's distance from Primary Site (a_3).

In order to handle the heavy data storages and server operations economically, Big Data requires a shift in cloud computing architecture (Villars et al., 2011). The flexible and scalable cloud computation technologies for DR applications allow data transactions over the Internet across multiple data centers being located in different geographic locations. Conventional architectures of cloud computing based DR systems rely on the tree-like network topologies. The capacities of different branches belonging to the tree-like network are typically oversubscribed. For such tree-like network topology, the extent of oversubscription is always raised to the level with very high hardware costs. Since the network connectivity cost calculation is based on the distance from the primary site to the DR site. The distance from the primary site to the DR site is still very important for the cloud computing based DR systems for Big Data applications.

Furthermore, in the Location and Infrastructure (A) dimension, the distance from the primary site (a_3) can be influenced by both natural disasters (a_1), manmade disasters (a_2), as well as detection and monitoring (a_6) (refer **Figure 2**). The rationality can be explained in this paragraph. For example, some data from the public sectors are highly confidential. If these public sectors adopt the cloud computing technology, more information security problems may happen. Such problems may have direct or indirect relations to their operations. Thus, many public sectors require the detection and monitoring mechanism (a_6), which not only to detecting the natural disasters (a_1) but also monitoring the hackers' attacks (a_2).

Managerial Implications from Continuity Management Perspective

This study combined the DNP and the VIKOR methods to identify the performance gaps of each criterion belonging to every target site. Because every site is constrained by limited finance, facility, and professional IT resources, the capability to prioritize factors to be improved is very critical for management teams to optimize the marginal effect. The improvement plan should be examined and conducted carefully to avoid any unexpected losses.

Improving Strategy for the Least Performing Site

For the Big Data environments, moving data is very hard to achieve. Moving the computing power to the data is more reasonable. The traditional DR systems provide one by one service. That is, the best-performing site is selected to provide DR services. So, the best-performing site should be improved.

However, for the modern geographic distributed DR sites for Big Data applications, the DR tasks are divided while multiple organizational goals shall be fulfilled. Each separate DR task is performed by the most suitable DR site from the dimensions of computation performance and telecommunications cost. Therefore, the DR ability of various data centers should be homogenized for the same applications. Thus, to provide distributed Big Data services, the Center X needs to improve the performance of site C to the level of site A, rather than improving the performance of site A to the aspired level (Villars et al., 2011).

Based on the VIKOR results, the DR Objectives (C) of site C is the least performing criterion. Furthermore, based on the analytic results being derived by the DEMATEL (refer **Figure 2**), the DR Objectives (C) dimension is influenced by both the IT System Availability (B) dimension and the Disaster Readiness Exercises (D) dimension.

For the ideal case, the availability of the IT systems belonging to the three sites of the Center X should be the same. However, after years of continuous purchases of new facilities and incremental upgrades of the DR systems, the IT System Availabilities (B) of the three sites is actually not the same. For the site C, the least performing criterion in the IT System Availability (B) dimension is the Backup System Architecture (b_3) (0.0450) (refer **Table 11**). The Backup System Architecture (b_3) is closely related to the data center's capability to take over from the primary site. If the Backup System Architecture (b_3) is a hot site, site C needs to be running immediately to meet RTO objectives. Site C needs to back up the original site of the organization, with full computer systems as well as near-complete backups of users' data. A real-time synchronization between the two sites may be used to completely mirror the data environment of the original site using the wide area network (WAN) links and specialized software.

In the Big Data era, modern IT systems need to cope with disasters effectively and efficiently (Sahebjamnia et al., 2015). The cloud based services (i.e. the Infrastructure as a Service, IaaS) can serve as a Backup System Architecture (b_3). Such cloud based services provide firms capabilities of data backups, failover of servers, and the ability to implement a secondary center which is far enough to allow for regional disaster recoveries.

Ideally, as a hot site comes up, the hot site initiates operations from a data processing perspective. The staff may be relocated to operations before personnel are moved to the hot site. The Emergency Operations Center (d_3) needs to take over all the activities and become the command center. According to the analytic results, Education and Training (d_1) and Emergency Operations Center (d_3) are key factors to be improved at first. In addition to the DR Procedure (e_3) and Testing and Exercises (c_3), Snedaker (2013) suggested that teams which are geographically dispersed should have various levels of accesses to the infrastructure, such as conference calls, video conferences, and e-mail. So, the performance of the least-performing site(s) can be enhanced.

Improving Strategy for the Best-Performing Site

According to **Table 11**, the Site A is the best-performing site based on the VIKOR results. In the dimension level, the VIKOR scores of the Operation Management (E) dimension of site A are significantly different from the others. The performance value versus the Top Manager's Support (e_4) criterion of site A more than the values of other two sites. The result is possibly due to the location of site A, which is located in headquarter of the Center X. All top managements of the Center X are located in site A. Therefore, in comparison with the other two sites, the Site A may provide better supports from the top management. This result is consistent with the previous studies by Herbane et al. (2004) who argued that the DR speed in responding to an incident depends on how quickly the organization identifies the incident and prepares for it. Organizational alertness and preparedness are easily improved if managers allocate resources and decide to implement backup plans and form crisis teams. If top management assigns responsibility to the IT department, it is able to quickly recognize potential risks and notify the crisis team (Herbane et al., 2004). If top management supports business continuity routine exercises throughout the whole organization, heightened awareness and commitment becomes part of the organizational culture for everyone (Alesi, 2008). This is consistent with the DEMATEL results being demonstrated in **Figure 2**. Both the IT System Availability (B) dimension and the DR Objectives (C) dimension are influenced by the Operation Management (E) dimension. Furthermore, top managers play the central role in Operation Management (E). In practice, if top managers respond quickly and support the second line engineer(s), the engineer(s) can quickly respond to the first-line engineer(s) and take over from the customer's primary site.

This research extends from the traditional DR site which focuses on IT infrastructure to the DR site in the Big Data era which focuses on operation management and disaster readiness exercises. Big Data environments consist of millions or even billions of small data files. Big Data constitutes challenges for the DR site's backup space, capacity, and recovery time. Moreover, higher data set volume increases the probability for data files and documents to contain inherently valuable and sensitive information. This also makes a more appealing target for hackers because Big Data information being stored is a potential goldmine for cybercriminals. Such crimes lead to an amplified information security technical and facility cost (Kshetri, 2014). For the whole emergency management system, operational system continuous improvement is also an important factor that influences the emergency management. To enhance and promote the effectiveness of emergency responses, the information security manager should ensure that each department being involved is aware of its duties and responsibilities. What's more, the centralized government authority is essential for the operation and execution of emergency responses. Such centralized authority assures the stability and the order of the management structure and thus, the performance of the overall emergency process, including planning, response, evacuation, relief, and reconstruction (Zhou et al., 2011) can be promoted.

Limitations and Future Research Possibilities

Three limitations exist for adopting the proposed MCDM framework of this study. First, although some exiting research has provided valuable insights for DR site selection, this research is the first attempt to select and evaluate the DR site for Big Data applications. The raw data was derived based on Taiwanese experts' opinions, which may be controversial. In this regard, future DR site selection researches can include studies of larger economies with more available experts. Second, this research uses the MCDM method to explore the factors being associated with DR sites. Since the DR site selection problem for Big Data is a highly pioneering and knowledge-intensive issue, the total number of available Taiwanese experts is less than 30, which does not fulfill the minimum number of samples for statistical calculations. Further research can use statistical analysis to measure the effect of the framework when more users adopt DR sites for Big Data. Third, the proposed framework of this research is focused on physical platforms of DR sites. However, how current big data environments or future cloud computing environments with virtual machines platforms (e.g., OpenStack) would influence the solution being proposed should be discussed further. Indeed, the virtual machine (VM) based platforms can influence the evaluation and selection of DR sites. In comparison with the VM-based platforms, the physical platforms move the physical servers to other locations along with the applications and data. Instead, VM-based platforms move applications and data to alternate DR sites over the network and automatically switch data handling to a standby system in the event of system compromise (i.e., automatic failover) through use of virtualization and cloud technologies. The VM-based DR solutions have some advantages over the physical platforms of DR sites. Compliance with different hardware devices of the VM-based platform may reduce the RTO (c_1) and RPO (c_2) (Khoshkholghi et al., 2014) and further improve resource management ability (e_3). However, the drawbacks of the VM-based platforms are significant, too. The VM-based platform needs to transmit the entire VM image and configuration during DR process. Thus, huge investments in network connectivity (b_4) are required. Additionally, license fees are required for using the commercial VM-based platforms (e.g., VMware, Microsoft cloud platform, Oracle stack, etc.). Management fees are required for the open source (e.g., VMware, Microsoft cloud platform, Oracle stack). In general, by considering both the pros and cons of the VM-based DR solutions, the physical platforms are more suitable. However, the future development of the VM-based DR solutions and whether such solutions can replace current physical platform based solutions is worth further study.

CONCLUSIONS

In the Big Data era, interest in effectively recovering massive data and systems in separate sites from disasters is emerging. How users can select DR sites for Big Data applications is very critical to ensure business continuity and investment optimization. The nature of the DRs for Big Data applications is indistinct and involves various considerations to be identified. Existing researches on the DR site selection problems have mainly focused on assessments of the probability and potential results of a disaster. Such research results cannot fulfill the needs for current and future Big Data applications. Thus, this research goes further and provides direct evaluation, selection, and improvement strategies for DR sites in general and the DR sites for Big Data applications in particular. This research defines an analytic framework for evaluating, selecting, and improving DR sites so as to reduce the gaps between the current status and the aspired level of all dimensions for selecting the best DR site. The hybrid MCDM methods based on the analytic framework consists of the DNP and the VIKOR methods. An empirical study based on evaluating and selecting a DR site for Big Data applications of an Asian high-performance computer center is used to demonstrate the feasibility of the proposed framework.

The analytic results provide IT managers with an understanding of the major concerns of DR site evaluation and selection for Big Data applications. The major issue raised in this paper is the focus on DR's fundamental values with respect to the business continuity plan. This is the key to all DR strategies in general. However, such issues are especially critical for Big Data applications. Although important, the RTO and RPO are costly, so the ideal DR site design for Big Data applications should consider both dimensions to meet the challenges of IT professionals. The content and application-aware schemes are the mainstays for business impact analysis evaluation. However, the cost of Big Data problems weighed on the axes of volume, velocity, variety, and variability is also of concerns when defining critical business. Another issue arising in the research is the more dominant roles of networks than those roles of the servers' or the storages' from the perspective of IT infrastructure. This research demonstrates how network costs are associated with security, data volume, data velocity, and data complexity. Therefore, the network bandwidth cost is a key factor for successful Big Data applications. Finally, this research also raises the importance of DR operation management. This research demonstrates how operation management influences IT facility availability since the network is more complex in Big Data. In summary, the analytic results can serve as a basis for IT managers' DR site selection for Big Data applications and, furthermore, can be used to optimize a DR site.

ACKNOWLEDGEMENTS

The authors thank the National Center for High-Performance Computing staff for inspiring conversations, suggestions, and feedback in preparing this article.

REFERENCES

Ablanedo-Rosas, J. H., Gao, H., Alidaee, B., & Teng, W.-Y. (2009). Allocation of emergency and recovery centres in Hidalgo, Mexico. *International Journal of Services Sciences, 2*(2), 206-218. doi:10.1504/IJSSCI.2009.024941

Al-Shaikh, R., Al-Hussain, Z., & Al-Sharidah, A. (2015). *Toward Building an IT Disaster Recovery Site for Oil and Gas Companies.* Paper presented at the Computational Intelligence, Communication Systems and Networks (CICSyN), 2015 7th International Conference on. doi:10.1109/CICSyN.2015.39

Alesi, P. (2008). Building enterprise-wide resilience by integrating business continuity capability into day-to-day business culture and technology. *Journal of Business Continuity & Emergency Planning, 2*(3), 214-220.

Alhazmi, O. H., & Malaiya, Y. K. (2012). *Assessing disaster recovery alternatives: On-site, colocation or cloud.* Paper presented at the Software Reliability Engineering Workshops (ISSREW), 2012 IEEE 23rd International Symposium on. doi:10.1109/ISSREW.2012.20

Altay, N., & Green, W. G. (2006). OR/MS research in disaster operations management. *European Journal of Operational Research, 175*(1), 475-493.

Andreolini, M., Colajanni, M., Pietri, M., & Tosi, S. (2015). Adaptive, scalable and reliable monitoring of big data on clouds. *Journal of Parallel and Distributed Computing, 79*, 67-79. doi:10.1016/j.jpdc.2014.08.007

Anthopoulos, L. G., Kostavara, E., & Pantouvakis, J.-P. (2013). An effective disaster recovery model for construction projects. *Procedia-Social and Behavioral Sciences, 74*, 21-30. doi:10.1016/j.sbspro.2013.03.026

Awasthi, A., Chauhan, S. S., & Goyal, S. K. (2011). A multi-criteria decision making approach for location planning for urban distribution centers under uncertainty. *Mathematical and Computer Modelling*, *53*(1), 98-109. doi:10.1016/j.mcm.2010.07.023

Bahrami, M., & Singhal, M. (2015). The role of cloud computing architecture in big data. *Information granularity, big data, and computational intelligence* (pp. 275-295): Springer. doi:10.1007/978-3-319-08254-7_13.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining, 1*(1), 3-17.

Balcik, B., & Beamon, B. M. (2008). Facility location in humanitarian relief. *International Journal of Logistics, 11*(2), 101-121. doi:10.1080/13675560701561789

Beggan, D. M. (2011). Disaster recovery considerations for academic institutions. *Disaster Prevention and Management: An International Journal, 20*(4), 413-422. doi:10.1108/09653561111161734

Berman, J. J. (2013). *Principles of big data: preparing, sharing, and analyzing complex information*: Newnes.

Bertrand, C. (2005). Business continuity and mission critical applications. *Network Security, 2005*(8), 9-11. doi:10.1016/S1353-4858(05)70269-5

Beyer, M. A., & Laney, D. (2012). *The importance of 'big data': a definition*. Stamford, CT: Gartner, 2014-2018.

Bowen, P., Hash, J., & Wilson, M. (2006). *SP 800-100. Information Security Handbook: A Guide for Managers*: NIST.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15*(5), 662-679. doi:10.1080/1369118X.2012.678878

Broder, J. F., & Tucker, E. (2011). *Risk analysis and the security survey*: Elsevier.

Brooks, C., Bedernjak, M., Juran, I., & Merryman, J. (2002). *Disaster Recovery Strategies*. IBM Redbook.

BS25999. (2007). Business Continuity Management-Part2: Specification Business Continuity Management., from http://www.25999.info/bs25999-2.htm.

Cegiela, R. (2006). *Selecting technology for disaster recovery.* Paper presented at the 2006 International Conference on Dependability of Computer Systems. 160 - 167. doi:10.1109/DEPCOS-RELCOMEX.2006.49

Cerullo, V., & Cerullo, M. J. (2004). Business continuity planning: a comprehensive approach. *Information Systems Management, 21*(3), 70-78. doi:10.1201/1078/44432.21.3.20040601/82480.11

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems, 63*, 67-80. doi:10.1016/j.dss.2013.08.008

Chang, V. (2015). Towards a Big Data system disaster recovery in a Private Cloud. *Ad Hoc Networks, 35*, 65-82. doi:10.1016/j.adhoc.2015.07.012

Chang, Y., Wilkinson, S., Potangaroa, R., & Seville, E. (2012). Managing resources in disaster recovery projects. *Engineering, Construction and Architectural Management, 19*(5), 557-580. doi:10.1108/09699981211259621

Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences, 275*, 314-347. doi:10.1016/j.ins.2014.01.015

Chen, R., Sharman, R., Rao, H. R., & Upadhyaya, S. J. (2008). Coordination in emergency response management. *Communications of the ACM, 51*(5), 66-73.

Claunch, C. (2004). *Management update: best practices in business continuity and disaster recovery*. Gartner Research.

Clitherow, D., Brookbanks, M., Clayton, N., & Spear, G. (2008). Combining high availability and disaster recovery solutions for critical IT environments. *IBM Systems Journal, 47*(4), 563. doi:10.1147/SJ.2008.5386509

Computing Research Association. (2015). *Data-Intensive Research in Education: Current Work and Next Steps.* Washington, D.C.: Computing Research Association.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science, 9*(3), 458-467. doi:10.1287/mnsc.9.3.458

Dawes, S. S., Cresswell, A. M., & Cahan, B. B. (2004). Learning from crisis lessons in human and information infrastructure from the world trade center response. *Social Science Computer Review, 22*(1), 52-66. doi:10.1177/0894439303259887

Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research? *Learning, Media and Technology, 38*(3), 237-240. doi:10.1080/17439884.2013.771783

Fan, Z.-P., Suo, W.-L., & Feng, B. (2012). Identifying risk factors of IT outsourcing using interdependent information: An extended DEMATEL method. *Expert systems with Applications, 39*(3), 3832-3840. doi:10.1016/j.eswa.2011.09.092

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5-6), 304-317. doi:10.1504/IJTEL.2012.051816.

Fontela, E., & Gabus, A. (1976). *The DEMATEL observer: Dematel 1976 Report*. Switzerland Geneva, Battelle Geneva Research Center.

Gallego, D., & Bueno, S. (2014). Exploring the application of the Delphi method as a forecasting tool in Information Systems and Technologies research. *Technology Analysis & Strategic Management, 26*(9), 987-999. doi:10.1080/09537325.2014.941348

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144. doi:10.1016/j.ijinfomgt.2014.10.007

Gantz, J., & Reinsel, D. (2011). *Extracting value from chaos*. IDC, 1142, 1-12.

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the Future, 2007*, 1-16.

Garlasu, D., Sandulescu, V., Halcu, I., Neculoiu, G., Grigoriu, O., Marinescu, M., & Marinescu, V. (2013). *A big data implementation based on Grid computing*. Paper presented at the Roedunet International Conference (RoEduNet), 2013 11th.1-4. doi:10.1109/RoEduNet.2013.6511732

Gibb, F., & Buchanan, S. (2006). A framework for business continuity management. *International journal of information management, 26*(2), 128-141. doi:10.1016/j.ijinfomgt.2005.11.008

Greenberg, A., Hamilton, J. R., Jain, N., Kandula, S., Kim, C., Lahiri, P., Sengupta, S. (2009). *VL2: a scalable and flexible data center network.* Paper presented at the ACM SIGCOMM computer communication review.

Gregory, P. H. (2011). *IT disaster recovery planning for dummies*: John Wiley & Sons.

Zeng, D., Gu, L., & Guo, S. (2015). *Cost minimization for big data processing in geo-distributed data centers.* In Cloud Networking for Big Data (pp. 59-78). Springer. doi:10.1007/978-3-319-24720-5_4

Han, H., Li, L., & Zhu, D. (2012). *Research and Implementation on Remote Disaster Recovery System.* Paper presented at the Computer Science & Service System (CSSS), 2012 International Conference on. 875 - 879. doi:10.1109/CSSS.2012.223

Hanaoka, S., & Qadir, F. M. (2005). *Logistics problems in recovery assistance of Indian Ocean earthquake and tsunami disaster.* Paper presented at the Scientific Forum on The Tsunami, Its Impact and Recovery.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98-115. doi:10.1016/j.is.2014.07.006

Herbane, B., Elliott, D., & Swartz, E. M. (2004). Business continuity management: time for a strategic role? *Long Range Planning, 37*(5), 435-457. doi:10.1016/j.lrp.2004.07.010

Hsieh, Y.-F., Lee, Y.-C., & Lin, S.-B. (2016). Rebuilding DEMATEL threshold value: an example of a food and beverage information system. *SpringerPlus, 5*(1), 1385. doi:10.1186/s40064-016-3083-7

Hsu, C.-W., Kuo, T.-C., Chen, S.-H., & Hu, A. H. (2013). Using DEMATEL to develop a carbon management model of supplier selection in green supply chain management. *Journal of cleaner production, 56*, 164-172. doi:10.1016/j.jclepro.2011.09.012

Hsu, C. W., Kuo, T. C., Shyu, G. S., & Chen, P. S. (2014). Low Carbon Supplier Selection in the Hotel Industry. *Sustainability, 6*(5), 2658-2684. doi:10.3390/su6052658

Huang, C.-Y., & Kao, Y.-S. (2015). UTAUT2 Based Predictions of Factors Influencing the Technology Acceptance of Phablets by DNP. *Mathematical Problems in Engineering, 501*, 603747. doi:10.1155/2015/603747

International Data Corporation (IDC). (2009). *Worldwide marketplace model and forecast*. Framingham, MA.

Järveläinen, J. (2013). IT incidents and business impacts: Validating a framework for continuity management in information systems. *International Journal of Information Management, 33*(3), 583-590. doi:10.1016/j.ijinfomgt.2013.03.001

Judson, J. (2012). *Disaster Recovery Best Practices - Templates, Documents and Examples of Disaster Recovery in the Public Domain PLUS access to content.theartofservice.com for downloading*: Emereo.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, 7-10 Jan. 2013). *Big Data: Issues and Challenges Moving Forward.* Paper presented at the System Sciences (HICSS), 2013 46th Hawaii International Conference on.

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing, 74*(7), 2561-2573. doi:10.1016/j.jpdc.2014.01.003

Katal, A., Wazid, M., & Goudar, R. (2013). *Big data: issues, challenges, tools and good practices.* Paper presented at the Contemporary Computing (IC3), 2013 Sixth International Conference on. doi:10.1109/IC3.2013.6612229

Kiernan, V. (2005a). Disasters may have wide ripple effects. *Chronicle of Higher Education, 52*(8), 32.

Kiernan, V. (2005b). Ready for the next Katrina. *Chronicle of Higher Education*, 52(8), A31-A33.

Kshetri, N. (2014). Big data' s impact on privacy, security and consumer welfare. *Telecommunications Policy, 38*(11), 1134-1145. doi:10.1016/j.telpol.2014.10.002

Khoshkholghi, M. A., Abdullah, A., Latip, R., Subramaniam, S., & Othman, M. (2014). Disaster recovery in cloud computing: A survey. *Computer and Information Science*, *7*(4), 39. doi:10.5539/cis.v7n4p39

Lam, J., Ng, K. K., Cheung, S. K., Wong, T. L., Li, K. C., & Wang, F. L. (2015). *Technology in Education. Technology-Mediated Proactive Learning: Second International Conference, ICTE 2015, Hong Kong, China, July 2-4, 2015, Revised Selected Papers* (Vol. 559): Springer. doi:10.1007/978-3-662-48978-9

Lampa, S., Dahlö, M., Olason, P. I., Hagberg, J., & Spjuth, O. (2013). Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. *Gigascience, 2*(1), 1. doi:10.1186/2047-217X-2-9

Lee, V. R. (November-December, 2013). The Quantified Self (QS) movement and some emerging opportunities for the educational technology field. *Educational Technology*, 39.

Li, C.-W., & Tzeng, G.-H. (2009). Identification of a threshold value for the DEMATEL method using the maximum mean de-entropy algorithm to find critical services provided by a semiconductor intellectual property mall. *Expert Systems with Applications, 36*(6), 9891-9898. doi:10.1016/j.eswa.2009.01.073

Liao, S., Wu, M.-J., Huang, C.-Y., Kao, Y.-S., & Lee, T.-H. (2014). Evaluating and Enhancing Three-Dimensional Printing Service Providers for Rapid Prototyping Using the DEMATEL Based Network Process and VIKOR. *Mathematical Problems in Engineering,* 1-16. doi:10.1155/2014/349348

Liebowitz, J. (2013). *Big data and business analytics*: CRC Press.

Lin, L., Shuang, W., Yifang, L., & Shouyang, W. (2014). A New Idea of Study on the Influence Factors of Companies' Debt Costs in the Big Data Era. *Procedia Computer Science, 31*, 532-541. doi:10.1016/j.procs.2014.05.299

Linstone, H. A., & Turoff, M. (2002). The Delphi Method. *Techniques and applications, 53*.

Liu, H. C., You, J. X., Zhen, L., & Fan, X. J. (2014). A novel hybrid multiple criteria decision making model for material selection with target-based criteria. *Materials & Design, 60*, 380-390. doi:10.1016/j.matdes.2014.03.071

Lu, M. T., Lin, S. W., & Tzeng, G. H. (2013). Improving RFID adoption in Taiwan's healthcare industry based on a DEMATEL technique with a hybrid MCDM model. [Article]. *Decision Support Systems, 56*, 259-269. doi:10.1016/j.dss.2013.06.006

Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society, 1*(2), doi:10.1177/2053951714541861

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.

Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal (KM&EL), 4*(3), 236-247.

Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think*: Houghton Mifflin Harcourt.

Mayer, B. W., Moss, J., & Dale, K. (2008). Disaster and preparedness: lessons from Hurricane Rita. *Journal of Contingencies and Crisis Management, 16*(1), 14-23. doi:10.1111/j.1468-5973.2008.00531.x

Minelli, M., Chambers, M., & Dhiraj, A. (2012). *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*: John Wiley & Sons.

Miyagawa, M. (2012). Joint distribution of distances to the first and the second nearest facilities. *Journal of geographical systems, 14*(2), 209-222. doi:10.1007/s10109-010-0143-3

Murry, J. W., & Hammons, J. O. (1995). Delphi: A versatile methodology for conducting qualitative research. *The Review of Higher Education, 18*(4), 423. doi:10.1353/rhe.1995.0008

Naranjo-Gil, D. (2009). Management information systems and strategic performances: The role of top team composition. *International Journal of Information Management, 29*(2), 104-110. doi:10.1016/j.ijinfomgt.2008.05.009

Opricovic, S. (1998). Multicriteria optimization of civil engineering systems. *Faculty of Civil Engineering, Belgrade, 2*(1), 5-21.

Opricovic, S., & Tzeng, G.-H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research, 156*(2), 445-455. doi:10.1016/S0377-2217(03)00020-1

Ou Yang, Y.-P., Shieh, H.-M., Leu, J.-D., & Tzeng, G.-H. (2008). A novel hybrid MCDM model combined with DEMATEL and ANP with applications. *International Journal of Operations Research, 5*(3), 160-168.

Papamitsiou, Z. K., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society, 17*(4), 49-64.

Pirkul, H., & Schilling, D. (1989). The capacitated maximal covering location problem with backup service. *Annals of Operations Research, 18*(1), 141-154. doi:10.1007/BF02097800

Rabbany, R., Takaffoli, M., & Zaïane, O. R. (2011). *Analyzing participation of students in online courses using social network analysis techniques.* Paper presented at the Proceedings of educational data mining.

Rodger, J. A., Bhatt, G., Chaudhary, P., Kline, G., & McCloy, W. (2015). The Impact of Business Expertise on Information System Data and Analytics Resilience (ISDAR) for Disaster Recovery and Business Continuity: An Exploratory Study. *Intelligent Information Management, 7*(04), 223. doi:10.4236/iim.2015.74017

Roebuck, K. (2012). *Business continuity and disaster recovery: High-impact Technology - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors*: Emereo Publishing.

Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 40*(6), 601-618.

Rothstein, P. J. (2007). *Disaster Recovery Testing: Exercising Your Contingency Plan (2007 Edition)*: Rothstein Associates Incorporated.

Saaty, T. L. (1999). *Decision making for leaders: the analytic hierarchy process for decisions in a complex world* (Vol. 2): RWS publications.

Sahebjamnia, N., Torabi, S. A., & Mansouri, S. A. (2015). Integrated business continuity and disaster recovery planning: Towards organizational resilience. *European Journal of Operational Research, 242*(1), 261-273. doi:10.1016/j.ejor.2014.09.055

Sembiring, J., & Siregar, M. I. H. (2013). A Decision Model for IT Risk Management on Disaster Recovery Center in an Enterprise Architecture Model. *Procedia Technology, 11*, 1142-1146. doi:10.1016/j.protcy.2013.12.306

Sengupta, S., & Annervaz, K. (2014). Multi-site data distribution for disaster recovery—A planning framework. *Future Generation Computer Systems, 41*, 53-64. DOI:10.1016/j.future.2014.07.007

Serrelis, E., & Alexandris, N. (2006). *Disaster Recovery Sites as a Tool of Managing Extreme Attacks.* Paper presented at the International Conference on Internet Surveillance and Protection (ICISP 06). doi:10.1109/ICISP.2006.9

Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE review, 46*(5), 30.

Smith, G. (2012). *Planning for Post-Disaster Recovery: A Review of the United States Disaster Assistance Framework*: Island Press.

Snedaker, S. (2013). *Business continuity and disaster recovery planning for IT professionals*: Newnes.

Sutton, J., & Tierney, K. (2006). *Disaster preparedness: concepts, guidance, and research.* Paper presented at the Fritz Institute Assessing Disaster Preparedness Conference.

Tammineedi, R. L. (2010). Business continuity management: A standards-based approach. *Information Security Journal: A Global Perspective, 19*(1), 36-50. doi:10.1080/19393550903551843

TechAdvisory.org. (2010). Disaster Recovery: An Increasingly Important Aspect of Your Business, from http://www.techadvisory.org/2010/05/disaster-recovery%E2%80%94an-increasingly-important-aspect-of-your-business/.

Thejendra, B. (2014). *Disaster Recovery and Business Continuity: A quick guide for small organisations and busy executives*: IT Governance Publishing.

Tian, W., & Zhao, Y. (2015). Big data technologies and cloud computing. *Optimized Cloud Resource Management and Scheduling, Morgan Kaufmann, Boston*, 17-49.

Tierney, K. J. (2007). From the margins to the mainstream? Disaster research at the crossroads. *Sociology, 33*(1), 503. doi:10.1146/annurev.soc.33.040406.131743

Tzeng, G.-H., & Huang, C.-Y. (2012). Combined DEMATEL technique with hybrid MCDM methods for creating the aspired intelligent global manufacturing & logistics systems. *Annals of Operations Research, 197*(1), 159-190. doi:10.1007/s10479-010-0829-4

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. *White Paper, IDC.*

Wallace, M., & Webber, L. (2010). *The disaster recovery handbook: A step-by-step plan to ensure business continuity and protect vital operations, facilities, and assets*: AMACOM Div American Mgmt Assn.

Wallace, M., Webber, L., & Webber, L. (2011). *The Disaster Recovery Handbook: A Step-by-step Plan to Ensure Business Continuity and Protect Vital Operations, Facilities, and Assets*: AMACOM.

Wang, N., & Pin, L. (2008). *Research on the Information Resources Management Center Construction in E-Government.* Paper presented at the 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing.

Wang, Y.-L., & Tzeng, G.-H. (2012). Brand marketing for creating brand value based on a MCDM model combining DEMATEL with ANP and VIKOR methods. *Expert systems with Applications, 39*(5), 5600-5615. doi:10.1016/j.eswa.2011.11.057

Webb, G. R., Tierney, K. J., & Dahlhamer, J. M. (2000). Businesses and disasters: Empirical patterns and unanswered questions. *Natural Hazards Review, 1*(2), 83-90. doi:10.1061/(ASCE)1527-6988(2000)1:2(83).

Whitson, G. (2003). Computer security: theory, process and management. *Journal of computing sciences in colleges, 18*(6), 57-66. doi:10.1108/09685220310500153

Wiboonrat, M. (2008). *An empirical IT contingency planning model for disaster recovery strategy selection.* Paper presented at the 2008 IEEE International Engineering Management Conference.

Yang, C.-L., Yuan, B. J., & Huang, C.-Y. (2015). Key Determinant derivations for information technology disaster recovery site selection by the multi-criterion decision making method. *Sustainability, 7*(5), 6149-6188. doi:10.3390/su7056149

Yang, Y.-P. O., Shieh, H.-M., & Tzeng, G.-H. (2013). A VIKOR technique based on DEMATEL and ANP for information security risk control assessment. *Information Sciences, 232*, 482-500. doi:10.1016/j.ins.2011.09.012

Yang, Y., Li, W., & Yuan, D. (2014). *Reliability Assurance of Big Data in the Cloud: Cost-effective Replication-based Storage*: Morgan Kaufmann.

Zhang, M. (2014). Who are interested in online science simulations? Tracking a trend of digital divide in Internet use. *Computers & Education, 76*, 205-214. doi:10.1016/j.compedu.2014.04.001

Zhang, N., Yan, Y., Xu, S., & Su, W. (2014). A distributed data storage and processing framework for next-generation residential distribution systems. *Electric Power Systems Research, 116*, 174-181. doi:10.1016/j.epsr.2014.06.005

Zhou, Q., Huang, W., & Zhang, Y. (2011). Identifying critical success factors in emergency management using a fuzzy DEMATEL method. *Safety Science, 49*(2), 243-252. doi:10.1016/j.ssci.2010.08.005

Zikopoulos, P. C., Eaton, C., DeRoos, D., Deutsch, T., & Lapis, G. (2012). *Understanding big data*. New York et al: McGraw-Hill.