

Measuring and evaluating data distribution strategies using an integrated approach of fuzzy based MOORA and AHP

Viral Gupta¹ · P. K. Kapur² · Deepak Kumar¹

Received: 20 September 2016 / Accepted: 5 January 2017 / Published online: 3 May 2017
© Society for Reliability and Safety (SRESA) 2017

Abstract Data distribution helps in achieving higher availability, reliability and increased performance of an enterprise application that is geographically dispersed. There are various data distribution strategies available and selection of a most suitable strategy for an enterprise is a multi-criteria decision-making problem. The selection problem of data distribution strategy is very complex because it depends on multiple conflicting criteria. It is essential to strike a right balance between conflicting criteria and select the most suitable data distribution strategy. This research paper presents a framework for measuring and selecting most suitable data distribution strategy using an integrated approach of fuzzy based multi-objective optimization on the basis of ratio analysis and analytic hierarchy process. A real world problem from an enterprise was undertaken to illustrate the results of the proposed methodology. In this paper, five data distribution strategies and six criteria were considered. The methodology used in this paper acts as an important tool for the practitioners to measure and evaluate the most suitable data distribution strategy in an enterprise and improve the overall quality.

Keywords Data distribution · Replication · Partitioning · Fuzzy MOORA · AHP · Quality

1 Introduction

Large enterprises having numerous departments, offices, users located in geographically dispersed locations require distributed database as centralized database cannot serve their needs. A distributed database is a single logical database that is physically distributed or spread across servers in multiple dispersed geographic locations; these servers are connected with each other on network. The fact that the database is distributed across multiple locations is not transparent to the users and the distributed database is still administered centrally. Some of the advantages that distributed databases offers are increased availability, reliability, performance, improved data sharing, and reduction in the data communication costs, ease of data distribution and autonomy of business units. There are numerous data distribution strategies that can be implemented for a large enterprise, some of the high-level strategies are data replication and data partitioning. It is an endeavor of data architects to select the most suitable data distribution strategy in an enterprise application. The selection of a most suitable data distribution strategy is very complex task as it depends on multiple conflicting criteria. There has been very little work done in the field of data distribution that can measure various criteria influencing the data distribution strategy and select the most suitable data distribution strategy quantitatively.

In this paper, we have proposed an integrated framework of fuzzy based MOORA and AHP to measure, evaluate and select the most suitable data distribution strategy. “Multi-objective optimization on the basis of ratio analysis”

✉ Viral Gupta
viralgupta@hotmail.com

P. K. Kapur
pkkapur1@gmail.com

Deepak Kumar
deepakgupta_du@rediffmail.com

¹ Amity Institute of Information Technology, Amity University, Noida, India

² Centre for Interdisciplinary Research, Amity University, Noida, India

(MOORA) is a technique to optimize two or more conflicting attributes. This technique is advancement over the traditional method of cost benefit analysis where the units of all the costs and benefits were required to be same. In MOORA, various attributes or criteria can have different units. The MOORA method was developed by Brauers and Zavadskas (2006), and has been used as one of the important tool for selecting the best alternative in the field of engineering and management. The MOORA technique considers the priorities of the conflicting criteria, which were calculated using analytic hierarchy process (AHP). The priority weights of the criteria were determined using pairwise comparison that was based on expert opinion. In this paper, an experiment was conducted in an enterprise application, total five data distribution strategy, six criteria were considered, and the most suitable data distribution strategy was selected using the proposed integrated framework. The priority weight of six conflicting criteria was determined using AHP and the most suitable data distribution strategy was selected using Fuzzy based MOORA.

The motivation behind this research work is the lack of an integrated model in the literature and need of such model, which can measure, evaluate and select the most suitable data distribution strategy in an environment. The remainder of this paper is structured in the following way. Section 2 describes the review of literature conducted during this paper, Sect. 3 presents the data distribution strategies and criteria, and Sect. 4 explains the proposed approach and experimental set up followed by the measures, outcomes, discussions and conclusion.

2 Literature review

The concepts related to data distribution strategies have been well established and discussed in the past (Ceri and Pelagatti 1984). There have been numerous articles related to the data distribution in literature. Ozsu and Valduriez (1991) studied distributed databases and analyzed the initial goals and promises of distributed databases; they evaluated how the commercial products were fulfilling these goals and promises. Some of the issues highlighted in their study are capability of distributed query processing, advanced transaction models, analysis of replication and its impact on distributed database, implementation strategies for distributed databases. The authors concluded that there are many technical problems related to distributed databases that await solution and need more research work. Rababaah and Hakimzadeh (2005) in their paper introduced the concepts and issues those were addressed by distributed databases. The authors evaluated various design alternatives for distributed databases and their

corresponding advantages like increased reliability, increased performance by reduction in the communication overhead, design, and data consistency. Various research works in the areas of query optimization, distribution, fragmentation were also examined. Reddy and Reddy (2012) proposed traditional approach followed to distribute the data in the cloud environment, which provided economical distribution of data those results in higher data availability and reliability. Various data fragmentation techniques were evaluated and the authors proposed a decision models for the cloud computing users that distributed the data across several service providers in an effective manner. Jin and Horng (2002) evaluated multiple approaches to data distribution and determined the best approach to obtain the good performance. Ye et al. (2001) in their paper discussed performance and scalability issues for distributed database servers, the authors proposed a generic data distribution strategy integrating user class information. Users were differentiated based on their access patterns. The distribution strategy proposed by authors consists of both partitioning and replication methods. Kamal and Murshed (2014) in their paper studied various challenges that are faced by architects and administrators while implementing data distribution strategy in a cloud environment. The authors discussed different replication and partitioning strategies available today and presented the potential way outs to overcome these challenges. Wang et al. (2008) developed two data distribution strategies to expedite the query response time through load balancing for query intensive data environment. Ashraf et al. (2013) studied the impact of fragmentation and distribution on the data retrieval time and selected the most appropriate fragmentation strategy based on the database architecture and selection patterns. It was concluded that the response time decreases as the strategy is changed from centralized to distributed databases.

Mazilu (2010) examined various database replication approaches and presented several advantages offered by them. The author also discussed the cases where different data replication technique can be implemented. Goel and Buyya (2006) presented a survey of replication algorithms used for different distributed systems and content management systems. The replication algorithms considered in this paper were distributed DBMS, P2P systems, data grid and WWW. These algorithms were evaluated on attributes namely reliability, performance, site autonomy, data control and heterogeneity. Srivastava et al. (2012) in their paper highlights basic concepts underlying the distributed database systems including transaction management and access control. The proposed approach of implementation of homogenous distributed database systems demonstrated the reduction in communication traffic and enhanced efficiency. Cheng et al. (2002) focused on the use of clustering

and explored the usage of genetic based clustering algorithm for data partitioning to achieve high data retrieval performance. The authors proposed three new genetic algorithm operators. Arres et al. (2015) presented a partitioning and placement approach to override the default policy and obtain improvement in query performance. This approach can be used by data warehousing solutions. Several other authors (Bhuyar et al. 2012; Chen et al. 2015; Kundakci 2016; Mazilu 2010) discussed partitioning and fragmentation strategies in distributed environment in large enterprises and presented how these strategies have addressed different technical and business challenges faced by the data architects and administrators.

MOORA is used in this paper to select the most appropriate data distribution strategy. Brauers and Zavadskas (2006) developed a new method namely multi-objective optimization on the basis of ratio analysis (MOORA) to rank and select the most suitable alternative. The method includes a matrix of responses of alternatives to objectives and the application of ratios. These dimensionless ratios lies between 0 and 1 are added in case of maximization and subtracted in case of minimization. The alternatives were ranked based on these combined ratios, which were considered as the best ratios. The authors illustrated the application of their proposed methodology in a transition company. Brauers and Zavadskas (2012) enhanced their earlier research work by joining MOORA with full multiplicative form for multiple objectives resulting into MULTI-MOORA, which eventually joined with DELPHI to give rise to the most robust approach for multi-objective optimization. Kundakci (2016) proposed an integrated approach of MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) and MULTI-MOORA (multi-objective optimization on the basis of ratio analysis) to rank the alternatives in the automobile selection problem in a marble company. MACBETH is used to determine the weights of the criteria. Görener et al. (2013) proposed an integrated approach of AHP and MOORA (multi-objective optimization on the basis of ratio analysis) to select the location of a bank in a banking industry. AHP is used to determine the weights of the criteria. Gadakh (2011) applied MOORA (multi-objective optimization on the basis of ratio analysis) to select optimal cutting parameters in manufacturing industry. Seema et al. (2014) applied FUZZY based MOORA for supplier selection problem in a chemical and biotechnical organization in India.

AHP is used in this paper to determine the priority weights for the criteria that influence the selection of data distribution strategy. Gupta et al. (2016) in their research paper presented a framework to model and measure the code smells in an enterprise using TISM and Two-way assessment. The priority weights for the code smells in an

enterprise were determined using AHP, where pairwise comparison was conducted based on expert opinion of stakeholders. Kapur et al. (2014) proposed a framework that leverages AHP to examine the implementation of ERP systems using AHP. The authors considered 10 critical success factors of ERP implementation. Pairwise comparison was conducted based on expert opinion and the priority weights for the critical success factors were determined. Authors were able to achieve the improvement of ERP implementation using the proposed methodology that was based on AHP.

After going through extensive literature review of data distribution strategies, MOORA and AHP frameworks, authors found that, there is a little amount of work done in research community on the selection of a data distribution strategy. However, there is lack of work that can quantitatively rank and select most suitable data distribution strategy. This motivated authors to take up the novel research work of measuring and selecting the most suitable data distribution strategy. Total five data distribution strategies and six criteria were considered in this study.

3 Data distribution strategies and the conflicting criteria

Today, numerous data distribution strategies are available that can be implemented in a large enterprise. The selection of a most suitable data distribution strategy depends on the conflicting criteria. In this paper, five strategies considered are centralized database, data replication with snapshot, data replication—near real time, horizontal partitioning and vertical partitioning. This section explains various data distribution strategies considered in this paper.

3.1 Centralized database

Centralized database is a concept where the database is stored in a single physical server. The entire user base needs access rights on the single server and accesses the data from central location. The advantages offered by centralized database are easy manageability, administration, better data integrity, security, and simplicity and data portability. The major disadvantages of centralized database are lack of fault tolerant systems, lack of high data availability, reliability and recoverability. There was over dependency on a single server, which causes a breakdown of the entire ecosystem in case there is any issue with the server like high traffic, faults or disaster. The centralized databases were more popular during the earlier days, when data distribution was not invented (Ozsu and Valduriez 1999).

3.2 Data replication with snapshot

Data replication is a fault tolerant technique used to create and maintain multiple copies of the same database at two or more sites. In Data replication, there are two types of databases namely master database and slave databases. Master database maintains the master copy of the database while the slave databases maintain the slave copies of the database. There can be two ways in which the contents of the master database are copied to the slave databases. These two ways are synchronized and non-synchronized replication. Synchronization is a way in which the data from one server is copied to other server immediately after an update on the first server is made. In synchronized replication, high level of data consistency is achieved. In the snapshot replication, the updates from all the replicated sites are sent to the master database; all the changes received from the replicated databases are collected in a snapshot log and sent to all replicated databases. The full or the incremental database snapshot from the master database is sent to the replicated databases.

3.3 Data replication: near real time

In the near real time replication scheme, the message for each completed transaction is broadcasted over the network to all other databases as soon as the data were updated at the originating database node. One way of broadcasting such messages over the network is database triggers. High level of data consistency is achieved in the near real time data replication. The reliability of the system is enhanced drastically with the near real time data replication.

3.4 Horizontal data partitioning

Partitioning in distributed database is defined as partitioning a single database table or relation into two or more partitions or fragments that can be stored at multiple sites and these partitions together represent the original table or the relation without any loss of information. In horizontal partitioning, a table is partitioned in such a way that some rows or the records of a table are stored at one site while other rows or records are stored at other sites. The database is designed in such a way that after horizontal partitioning the rows stored at a particular site represents the data most frequently used by the corresponding site (Navathe et al. 1995). The horizontal data partitioning provides data efficiency, local optimization, data security and greater ease of querying data.

3.5 Vertical data partitioning

In vertical partitioning, a table is partitioned in such a way that some columns of a table are stored at one site while other

columns are stored at other sites. The database is designed in such a way that after vertical partitioning the columns stored at a particular site represents the data most frequently used by that site (Navathe et al. 1984). The vertical data partitioning provides data efficiency, local optimization, data security and greater ease of querying data.

The six criteria considered in this paper, that influences the data distribution strategy are reliability, expandability, communication overhead, manageability, data consistency and overall costs. Reliability (Kapur et al. 2011) is a quality attribute that implies the probability of the defect free operation of a software application for a specified period in a specified environment. In case one server is down or busy then the other server should be able to serve the user request and hence provide increased reliability (Kapur et al. 1999) of the software application. Expandability (Somerville 2001) of the application means the ability of the system to incorporate more additions to the existing structure. Manageability (Somerville 2001) is another quality attribute that defines the ability of the system to be managed from different perspective. Manageability includes concerns like ease of data query, data coordination, and capability of handing data collisions. Data consistency (Somerville 2001) refers to the state where the data values remain same for all the instances of the data. This is one of the critical and most desirable requirement in data distribution, where it is required that the data values should be same across all the replicated sites. Communication overhead includes the cost of communication of data across multiple instances of databases. Finally, overall cost is another important criterion that influences the data distribution strategy in an enterprise. Overall costs include the hardware and software costs to set up the data distribution strategy. These six criteria are conflicting as data architects targets to maximize reliability, expandability, manageability and data consistency while reducing the communication overhead and the overall costs.

4 Proposed approach and experimental set up

This paper uses integrated framework comprising of FUZZY based MOORA and AHP to measure the criteria and select the most suitable data distribution strategy in an enterprise application. An enterprise application integrates disparate information technology systems to perform end-to-end business processes in an organization (Kumar et al. 2015). An experiment is conducted on an enterprise application also known as retail banking transaction system (RBTS). RBTS is a banking application where users across the world are required to access the application for making banking transactions. The practitioners and the data architects of the project are tasked to evaluate various

distribution strategies and select and implement the most suitable data distribution strategy. The selection of the data distribution strategy is a multi-criteria decision-making (MCDM) problem that is dependent on various conflicting criteria. The five data strategies considered in this paper are centralized database, data replication with snapshot, data replication—near real time, horizontal partitioning and vertical partitioning. The six criteria considered in this paper are reliability, expandability, communication overhead, manageability, data consistency and costs. These strategies and criteria are selected from the literature and are consulted from the industry experts.

4.1 Fuzzy decision matrix

In the fuzzy based MOORA technique, the first step is to establish the fuzzy decision matrix, in which responses for the alternatives on the criteria are obtained. The responses are taken from the key decision-makers that have in-depth experience of data distribution strategies in an enterprise. Total five experts, who will act as decision-makers were selected. Each alternative is evaluated based on every criterion. In the below decision matrix, $x_{ij}^a, x_{ij}^b, x_{ij}^c$ represents lower, middle and higher values for the triangular membership function for the i th alternative on the j th criteria.

$$X = \begin{bmatrix} [x_{11}^a, x_{11}^b, x_{11}^c] & [x_{12}^a, x_{12}^b, x_{12}^c] & [x_{13}^a, x_{13}^b, x_{13}^c] & \dots & [x_{1n}^a, x_{1n}^b, x_{1n}^c] \\ [x_{21}^a, x_{21}^b, x_{21}^c] & [x_{22}^a, x_{22}^b, x_{22}^c] & [x_{23}^a, x_{23}^b, x_{23}^c] & \dots & [x_{2n}^a, x_{2n}^b, x_{2n}^c] \\ [x_{31}^a, x_{31}^b, x_{31}^c] & [x_{32}^a, x_{32}^b, x_{32}^c] & [x_{33}^a, x_{33}^b, x_{33}^c] & \dots & [x_{3n}^a, x_{3n}^b, x_{3n}^c] \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ [x_{m1}^a, x_{m1}^b, x_{m1}^c] & [x_{m2}^a, x_{m2}^b, x_{m2}^c] & [x_{m3}^a, x_{m3}^b, x_{m3}^c] & \dots & [x_{mn}^a, x_{mn}^b, x_{mn}^c] \end{bmatrix} \tag{2}$$

AHP is used to measure the six criteria and assign priority weights to these criteria, these weights are then used by FUZZY based MOORA. AHP is an analytic tool used to address complex decision-making problems by converting the qualitative values into quantitative values. MOORA, which is developed by Brauers, is a technique to optimize two or more conflicting attributes (or criteria) and helps in ranking or selection of an alternative.

FUZZY set theory was developed by Zadeh (1965) and has been very popularly used in the areas where there is an element if uncertainty and it is difficult to define the system precisely. The theory helps in modeling the system qualitatively and quantitatively when there vagueness, ambiguity and uncertainty exists. There can be two types of fuzzy numbers namely triangular fuzzy number and trapezoidal fuzzy numbers. In this paper, triangular fuzzy numbers are considered. Let X be the universe of discourse, a fuzzy sub-set A is represented by a membership function $\mu_A(x)$ that maps each element x in X to the real number in the interval $[0, 1]$. A triangular fuzzy number is composed of three numbers (a, b, c) whose membership function $\mu_A(x)$ is given by:

$$\begin{aligned} \mu_A(x) &= \frac{x - a}{x - b}, & a \leq x \leq b \\ &= \frac{x - c}{b - c}, & b \leq x \leq c \\ &= 0, & \text{others.} \end{aligned} \tag{1}$$

Following 6-step methodology is adopted to achieve the objective of the paper.

4.2 Normalized fuzzy decision matrix

In the next step, the fuzzy decision matrix created in step 1 is normalized using vector normalized technique. The element values for the normalized fuzzy decision matrix are calculated as per below equations:

$$t_{ij}^a = \frac{x_{ij}^a}{\sqrt{\sum_{l=1}^m [(x_{ij}^a)^2 + (x_{ij}^b)^2 + (x_{ij}^c)^2]}} \tag{3}$$

$$t_{ij}^b = \frac{x_{ij}^b}{\sqrt{\sum_{l=1}^m [(x_{ij}^a)^2 + (x_{ij}^b)^2 + (x_{ij}^c)^2]}} \tag{4}$$

$$t_{ij}^c = \frac{x_{ij}^c}{\sqrt{\sum_{l=1}^m [(x_{ij}^a)^2 + (x_{ij}^b)^2 + (x_{ij}^c)^2]}} \tag{5}$$

4.3 Determination of priority weights for the criteria

In this step, the priority weight for the criteria is determined using AHP. In this technique, the criteria are organized and pairwise comparison of these criteria is conducted using the expert opinion of the stakeholders. Pairwise comparison is a mechanism in which, all the attributes/criteria are compared among each other, two at

a time. During the pairwise comparison, experts compare two criteria at a time and their qualitative judgement is converted into quantitative measure using Saaty's 1–9 Scale (Saaty 2008). The results of the pairwise comparison are captured in a matrix structure known as judgement matrix. Once the numerical values are assigned in the judgement matrix, the initial matrix is normalized and the priority weight (w) is calculated. The score of a_{ij} in the pairwise judgement matrix represents the relative importance of the element in the (i) row over the element in the (j) column. A sample judgement matrix is depicted below.

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & 1 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & 1 & \cdots & a_{3n} \\ a_{41} & a_{42} & a_{43} & \cdots & a_{4n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 1 \end{bmatrix}$$

Multiple algorithms are available to calculate the priority weight (w). Following algorithm is used in this paper to calculate the priority weight (w). In this algorithm, “ J ” is the column number while “ I ” is the row number.

$$W_i = \frac{\sum_{i=1}^I \left(\frac{a_{ij}}{\sum_{j=1}^J a_{ij}} \right)}{J} \quad (6)$$

4.4 Weighted normalized fuzzy decision matrix

The priority weights obtained in the previous step using AHP is multiplied by each element of the normalized fuzzy decision matrix to derive weighted normalized fuzzy decision matrix.

$$v_{ij}^a = W_i t_{ij}^a \quad (7)$$

$$v_{ij}^b = W_i t_{ij}^b \quad (8)$$

$$v_{ij}^c = W_i t_{ij}^c \quad (9)$$

4.5 Overall ratings for the beneficial and non-beneficial criteria

In this step, the overall ratings of the beneficial and non-beneficial criteria are calculated for each data distribution strategy. The overall rating for a data distribution strategy for the beneficial criteria for lower, middle and upper values for the triangular function can be calculated as per below equations.

$$s_i^{+a} = \sum_{j=1}^n v_{ij}^a, \quad \text{where } j \text{ belongs to the beneficial criteria} \quad (10)$$

$$s_i^{+b} = \sum_{j=1}^n v_{ij}^b, \quad \text{where } j \text{ belongs to the beneficial criteria} \quad (11)$$

$$s_i^{+c} = \sum_{j=1}^n v_{ij}^c, \quad \text{where } j \text{ belongs to the beneficial criteria} \quad (12)$$

$$s_i^{-a} = \sum_{j=1}^n v_{ij}^a, \quad \text{where } j \text{ belongs to the non-beneficial criteria} \quad (13)$$

$$s_i^{-b} = \sum_{j=1}^n v_{ij}^b, \quad \text{where } j \text{ belongs to the non-beneficial criteria} \quad (14)$$

$$s_i^{-c} = \sum_{j=1}^n v_{ij}^c, \quad \text{where } j \text{ belongs to the non-beneficial criteria} \quad (15)$$

4.6 Overall performance index for each alternative

In this step, the overall performance index (S_i) for each alternative is calculated as per below equation, which helps in achieving the de-fuzzified numbers.

$$S_i(s_i^+, s_i^-) = \sqrt{\frac{1}{3} \left[(s_i^{+a} - s_i^{-a})^2 + (s_i^{+b} - s_i^{-b})^2 + (s_i^{+c} - s_i^{-c})^2 \right]} \quad (16)$$

The data distribution strategy having the highest overall performance index is considered as the most suitable data distribution strategy.

5 Measures, outcomes and discussions

This sections illustrates the outcome of the methodology applied in the enterprise application RBTS. As the first step, five data distribution strategies (DS1 = centralized database, DS2 = data replication with snapshot, DS3 = data replication—near real time, DS4 = horizontal partitioning and DS5 = vertical partitioning) were evaluated against each of the six criteria (R = reliability, E = expandability, CO = communication overhead, M = manageability, D = data consistency and C = costs). The resultant fuzzy decision matrix is illustrated in Table 1 below.

The fuzzy decision matrix is then normalized using the Eqs. (3), (4) and (5). The normalized fuzzy decision matrix is illustrated in Table 2.

Table 1 Fuzzy decision matrix

	<i>R</i>	<i>E</i>	CO	<i>M</i>	<i>D</i>	<i>C</i>
DS1	(0.2, 0.4, 0.6)	(0.2, 0.2, 0.4)	(0.8, 0.8, 1)	(0.8, 1, 1)	(0.8, 0.8, 1)	(0.2, 0.4, 0.6)
DS2	(0.4, 0.6, 0.8)	(0.2, 0.4, 0.6)	(0.2, 0.4, 0.6)	(0.6, 0.8, 0.8)	(0.6, 0.8, 0.8)	(0.4, 0.6, 0.8)
DS3	(0.8, 0.8, 1)	(0.6, 0.8, 1)	(0.4, 0.6, 0.8)	(0.6, 0.8, 0.8)	(0.6, 0.8, 1)	(0.8, 1, 1)
DS4	(0.8, 0.8, 1)	(0.8, 0.8, 1)	(0.2, 0.4, 0.4)	(0.4, 0.6, 0.8)	(0.4, 0.6, 0.8)	(0.4, 0.6, 0.8)
DS5	(0.4, 0.6, 0.8)	(0.8, 0.8, 1)	(0.2, 0.4, 0.4)	(0.4, 0.6, 0.8)	(0.4, 0.6, 0.8)	(0.4, 0.6, 0.8)

Table 2 Normalized fuzzy decision matrix

	<i>R</i>	<i>E</i>	CO	<i>M</i>	<i>D</i>	<i>C</i>
DS1	(0.073, 0.147, 0.220)	(0.074, 0.074, 0.147)	(0.368, 0.368, 0.460)	(0.279, 0.348, 0.348)	(0.279, 0.279, 0.348)	(0.077, 0.155, 0.232)
DS2	(0.147, 0.220, 0.293)	(0.074, 0.147, 0.221)	(0.092, 0.184, 0.276)	(0.209, 0.279, 0.279)	(0.209, 0.279, 0.279)	(0.155, 0.232, 0.310)
DS3	(0.293, 0.293, 0.367)	(0.221, 0.295, 0.369)	(0.184, 0.276, 0.368)	(0.209, 0.279, 0.279)	(0.209, 0.279, 0.348)	(0.310, 0.387, 0.387)
DS4	(0.293, 0.293, 0.367)	(0.295, 0.295, 0.369)	(0.092, 0.184, 0.184)	(0.139, 0.209, 0.279)	(0.139, 0.209, 0.279)	(0.155, 0.232, 0.310)
DS5	(0.147, 0.220, 0.293)	(0.295, 0.295, 0.369)	(0.092, 0.184, 0.184)	(0.139, 0.209, 0.279)	(0.139, 0.209, 0.279)	(0.155, 0.232, 0.310)

Table 3 Criteria weights-using AHP

Criteria	<i>R</i>	<i>E</i>	CO	<i>M</i>	<i>D</i>	<i>C</i>	Weight
Reliability	0.2069	0.2857	0.4444	0.2105	0.0541	0.1905	0.2320
Expandability	0.0690	0.0952	0.1111	0.0526	0.0541	0.1905	0.0954
Communication overhead	0.1034	0.1905	0.2222	0.3158	0.4865	0.2857	0.2674
Manageability	0.1034	0.1905	0.0741	0.1053	0.1622	0.0476	0.1138
Data consistency	0.4138	0.1905	0.0741	0.1053	0.1622	0.1905	0.1894
Costs	0.1034	0.0476	0.0741	0.2105	0.0811	0.0952	0.1020

The weights for various criteria were calculated using AHP. Qualitative judgment values were obtained by the expert opinion. The qualitative values were converted into quantitative values and the final priority weights for the six criteria were obtained using Eq. (6). The priority weights are illustrated in Table 3 below.

The elements of the normalized fuzzy decision matrix were multiplied by the weights of the criteria determined in table-3 and the weighted normalized fuzzy decision matrix is obtained. The resultant weighted normalized fuzzy decision matrix is illustrated in Table 4.

Finally, overall performance ratings for each data distribution strategy were calculated using Eqs. (10)–(16). The beneficial criteria were reliability, expandability, manageability and

data consistency while the non-beneficial criteria were communication overhead and overall costs. The beneficial criteria implies that the values related to these criteria needs to be maximized or added and the non-beneficial criteria implies that the values related to these criteria needs to be minimized hence these values were subtracted. The overall performance ratings for all the five data distribution strategy and the rankings of these data distribution strategies are illustrated in Table 5.

It is found that an integrated approach of fuzzy based MOORA and AHP has successfully measured the criteria influencing the selection problem of data distribution strategies, have ranked, and selected the most suitable data distribution strategy. In the first step, the experts from the industry had evaluated all the 5 data distribution strategies

Table 4 Weighted normalized fuzzy decision matrix

	<i>R</i>	<i>E</i>	CO	<i>M</i>	<i>D</i>	<i>C</i>
DS1	(0.017, 0.034, 0.051)	(0.007, 0.007, 0.014)	(0.098, 0.098, 0.123)	(0.032, 0.040, 0.040)	(0.053, 0.053, 0.066)	(0.008, 0.016, 0.024)
DS2	(0.034, 0.051, 0.068)	(0.007, 0.014, 0.021)	(0.025, 0.049, 0.074)	(0.024, 0.032, 0.032)	(0.040, 0.053, 0.053)	(0.016, 0.024, 0.032)
DS3	(0.068, 0.068, 0.085)	(0.021, 0.028, 0.035)	(0.049, 0.074, 0.098)	(0.024, 0.032, 0.032)	(0.040, 0.053, 0.066)	(0.032, 0.039, 0.039)
DS4	(0.068, 0.068, 0.085)	(0.028, 0.028, 0.035)	(0.025, 0.049, 0.049)	(0.016, 0.024, 0.032)	(0.026, 0.040, 0.053)	(0.016, 0.024, 0.032)
DS5	(0.034, 0.051, 0.068)	(0.028, 0.028, 0.035)	(0.025, 0.049, 0.049)	(0.016, 0.024, 0.032)	(0.026, 0.040, 0.053)	(0.016, 0.024, 0.032)

Table 5 Overall performance ratings

	Beneficial criteria (s_i^+)			Non-beneficial criteria (s_i^-)			Overall (S_i)	Rank
DS1	0.109	0.133	0.171	0.106	0.114	0.147	0.0178	5
DS2	0.104	0.150	0.174	0.040	0.073	0.105	0.0699	4
DS3	0.153	0.181	0.218	0.081	0.113	0.138	0.0732	3
DS4	0.138	0.160	0.205	0.040	0.073	0.081	0.1041	1
DS5	0.104	0.143	0.188	0.040	0.073	0.081	0.0824	2

against the 6 criteria using the fuzzy function. The data obtained was captured in the fuzzy decision matrix, which was then normalized to obtain the normalized fuzzy decision matrix. The integrated approach measured the 6 criteria using AHP. In AHP, all the 6 criteria were undergone pairwise comparison using expert opinion. The expert opinion was converted into quantitative data and the priority weights for each criteria was determined. It was found that the maximum weight was allocated to communication overhead, followed by reliability, data consistency, manageability, costs and expandability. The weights of the criteria were utilized in the FUZZY based MOORA technique and the weighted normalized fuzzy decision matrix was obtained. Finally overall performance ratings of different data distribution strategies were determined after applying the equations of MOORA on the beneficial and non-beneficial criteria. The beneficial criteria considered were reliability, expandability, manageability and data consistency while the non-beneficial criteria considered were communication overhead and overall costs. The values of the beneficial criteria were added and the values of the non-beneficial criteria were subtracted from the overall performance score. Based on the overall performance score, the final ranking of the distribution strategies were derived and it was found that horizontal partitioning was the most suitable data distribution strategy followed by vertical partitioning, data replication with near real time, data replication with snapshot and centralized database. The results obtained from the application of the methodology were discussed with the practitioners, data architects and administrators of the RBTS project and they accepted the results by considering the horizontal partitioning in their project. The proposed methodology is a very effective tool for the practitioners to measure the criteria and select the most suitable data distribution strategy in an enterprise.

6 Conclusions

Data distribution is a very essential aspect of large enterprises that enables geographically dispersed departments, customers, users, employees to access the database in an effective manner. There are various data distribution strategies available that provides higher availability,

reliability and increased performance of an enterprise application. The selection of the most suitable data distribution strategy is very complicated, multi-criteria decision-making problem as it depends on conflicting criteria. There is a lack of work in the literature that can quantitatively measure the criteria and select the most suitable data distribution strategy in an enterprise. In this paper an integrated approach of FUZZY based MOORA and AHP is applied in an enterprise application RBTS, where the 6 criteria (reliability, expandability, communication overhead, manageability, data consistency and costs) were measured and the most suitable data distribution strategy was determined out of 5 available strategies (centralized database, data replication with snapshot, data replication—near real time, horizontal partitioning and vertical partitioning). The priority weights for the criteria were obtained using AHP and it was found that communication overhead received maximum weight followed by reliability. These weights were considered in FUZZY based MOORA and it was derived that the most suitable data distribution strategy was horizontal partitioning followed by vertical partitioning. This research has contributed to the field of data distribution by proposing an integrated methodology that proves out to be an important tool for the practitioners to measure the criteria and determine the most suitable data distribution strategy in an enterprise.

Acknowledgements Authors of this paper are sincerely grateful to The Founder President of Amity University, Dr. Ashok K. Chauhan, who has overwhelmingly shown his keen interest in fostering research in the Amity Universe and has always been a motivation for achieving greater triumphs.

References

- Arres B, Kabachi N, Boussaid O (2015) A data pre-partitioning and distribution optimization approach for distributed data warehouses. In: Proceedings of the international conference on parallel and distributed processing techniques and applications, pp 454–461
- Ashraf I, Khokhar AS, Lundberg L (2013) Impact of data distribution on response time in telecom databases. *Int J Eng Res Appl* 3(6):404–412
- Bhuyar PR, Gawande AD, Deshmukh AB (2012) Horizontal fragmentation technique in distributed database. *Int J Sci Res Publ* 2(5):1–7

- Brauers WKM, Zavadskas EK (2006) The MOORA method and its application to privatization in a transition economy. *Control Cybern* 35(2):445–469
- Brauers WKM, Zavadskas EK (2012) Robustness of MULTI-MOORA: a method for multi-objective optimization. *Informatika* 23(1):1–25
- Ceri S, Pelagatti G (1984) *Distributed databases principles and system*, 1st edn. McGraw-Hill, New York
- Chen K, Zhou Y, Cao Y (2015) Online data partitioning in distributed database systems. In: *EDBT*, pp 1–12
- Cheng CH, Lee WK, Wong KF (2002) A genetic algorithm-based clustering approach for database partitioning. *IEEE Trans Syst Man Cybern* 32(3):215–230
- Gadakh VS (2011) Application of MOORA method for parametric optimization of milling process. *Int J Appl Eng Res* 1(4):743–758
- Goel S, Buyya R (2006) Data replication strategies in wide area distributed systems. *Enterprise service computing: from concept to deployment*, p 17
- Görener A, Dinçer H, Hacıoğlu U (2013) Application of multi-objective optimization on the basis of ratio analysis (MOORA) method for bank branch location selection. *Int J Financ Bank Stud* 2(2):41–52
- Gupta V, Kapur PK, Kumar D (2016) Modelling and measuring code smells in enterprise applications using TISM and two-way assessment. *Int J Syst Assur Eng Manag* 7(3):332–340
- Jin MH, Horng JT (2002) Evaluating the performance of a distributed database of repetitive elements in complete genomes. *J Inform Sci Eng* 18:533–539
- Kamal JM, Murshed M (2014) *Distributed database management systems: architectural design choices for the cloud*. In: *Cloud computing*. Springer International Publishing, pp 23–50
- Kapur PK, Garg RB, Kumar S (1999) *Contributions to hardware and software reliability*. World Scientific, Singapore
- Kapur PK, Pham H, Gupta A, Jha PC (2011) *Software reliability assessment with OR applications*. Springer, London
- Kapur PK, Nagpal S, Khatri SK (2014) Critical success factor utility based tool for ERP health assessment—a general framework. *Int J Syst Assur Eng Manag* 5(2):133–148
- Kumar D, Gupta V, Kapur PK (2015) Assessment of quality factors in enterprise application integration. In: *4th international conference on reliability, infocom technologies and optimization (ICRITO)*, IEEE, Noida, India, pp 1–6
- Kundakci N (2016) Combined multi-criteria decision making approach based on MACBETH and MULTI-MOORA methods. *J Oper Res Stat Econ Manag Inform Syst* 4(1):17–26
- Mazilu MC (2010) Database replication. *Database Syst J* 1(2):33–38
- Navathe S, Ceri S, Wiederhold G, Dou J (1984) Vertical partitioning algorithms for database design. *ACM Trans Database Syst (TODS)* 9(4):680–710
- Navathe S, Karlapalem K, Ra M (1995) A mixed fragmentation methodology for initial distributed database design. *J Comput Softw Eng* 3(4):395–426
- Ozsu MT, Valduriez P (1991) *Distributed database systems: where are we now?* IEEE Comput 24(8):68–78
- Ozsu MT, Valduriez P (1999) *Principles of distributed database systems*, 2nd edn. Prentice-Hall, New Jersey
- Rababaah H, Hakimzadeh DH (2005) *Distributed databases fundamentals and research*. Advanced Database-B561. Springer
- Reddy BA, Reddy PR (2012) Effective data distribution techniques for multi-cloud storage in cloud computing. *Int J Eng Res Appl* 2(5):1130–1134
- Saaty TL (2008) Decision making with the analytic hierarchy process. *Int J Ser Sci* 1(1):83–98
- Seema, Kaur R, Kumar D (2014) Designing a mathematical model using fuzzy based MOORA method for supplier selection. *Int J Adv Eng Technol* 5(1):16–24
- Somerville I (2001) *Software engineering*, 6th edn. Addison-Wesley, Boston
- Srivastava A, Shankar U, Tiwari SK (2012) Transaction management in homogenous distributed real-time replicated database systems. *Int J Adv Res Comput Sci Softw Eng* 2(6):190–196
- Wang T, Yang B, Gao J, Yang D (2008) Effective data distribution and reallocation strategies for fast query response in distributed query-intensive data environments. In: *Asia-Pacific web conference*, Springer, Berlin, pp 548–559
- Ye H, Kerhervé B, Bochmann GV, Bourne D (2001) Data distribution strategies for providing database scalability in e-commerce applications. In: *Third international workshop on advanced issues of E-commerce and web-based information systems, WECWIS 2001*. IEEE, pp 86–88
- Zadeh LA (1965) Fuzzy sets. *Int J Inform Control* 8:338–353