Integrating Dependability Concepts while Selecting Cloud Service for Big Data

Fatima Ezzahra Mdarbi
RITM Laboratory, EST, ENSEM, Hassan II University Casablanca, Morocco
Email: Fati.mdarbi@gmail.com

Nadia Afifi
RITM Laboratory, EST, Hassan II University Casablanca, Morocco
Email: nafifi@est-uh2c.ac.ma

Imane Hilal
School of Information Sciences, Rabat, Morocco
Email: ihilal@esi.ac.ma

***Abstract*—The increased use of the Cloud to process Big Data has impacted the growth of its services. These various services have made tedious the choice of the suitable Cloud. Moreover, the sources of Big Data generate increasingly voluminous data with various levels of confidence. If the data are not available, reliable, secure, and maintainable, the resulting decisions are likely to be biased. Therefore, to fully and correctly use these Big Data, we should take into consideration their dependability requirements in the Cloud Service selection process. In this paper, we present a new approach for selecting Cloud Services based on a multi-criteria decision method called Fuzzy Analytic Hierarchy Process. Our contribution aims to integrate dependability concepts into the process of selecting the cloud service adapted to Big Data. This approach represents a good interest for end-users, by providing them the ability to specify their priorities in terms of dependability requirements of Big Data. To validate our approach, we implement a system to automate the Cloud selection process. Then we applied our proposal to a case study to demonstrate its feasibility and efficacy.**

***Index Terms*—big data, cloud services, dependability attributes, fuzzy analytical hierarchy process**

I. Introduction

Nowadays, Big Data is one of the most discussed issues in the computer technology field. Many organizations are interested in Big Data and their perspective since these data are valuable and contribute to the development of these companies. Big Data is defined as a large quantity of data set that is complex enough to be processed using traditional tools. Big Data is characterized by 5 characteristics known by the 5V: Volume, Velocity, Variety, Veracity, and Value [1]. Big Data management is confronted with many challenges, including: (i) The inadequacy of using classic Database Management Systems (DBMS) to process such data as they are not designed to store, manage and analyze Big Data. This is why Cloud Computing was adopted to manage Big Data [2], and (ii) Data quality and security issues, as well as their availability and reliability, are among the major concerns of organizations as they are faced with heterogeneous infrastructures and diverse users [3]. Cloud Computing represents an important economic and strategic stake. It introduced a major change in both the Big Data project's deployment and exploitation of IT resources. Indeed, organizations have no longer to invest in IT resources that require expensive and complicated internal management, since Cloud Computing offers various services and resources available. The diversity of cloud services makes it very difficult for users to choose the most appropriate service. Therefore, an efficient and accurate decision-making approach is highly recommended to select the cloud service that meets users' requirements. The cloud service selection problem is considered a multi-criteria decision-making problem [4]. Among the most widely used methods for selection, we find (i) Analytic Hierarchy Process (AHP), and (ii) Fuzzy Analytic Hierarchy Process (FAHP). AHP fragments a complex multi-criteria problem into a hierarchy, and it is based on a pairwise comparison of the different criteria [5]. FAHP is an extension of AHP, it is based on human evaluation and captures imprecision in ordinal judgments. It determines weights with precision [6] (p. 201).

In this paper, we present an analytical study of dependability attributes (availability, reliability, confidentiality, integrity, maintainability) and the 5V (Volume, Velocity, Variety, Veracity, Value) of Big Data to define the points of convergence and to determine the guidelines to be respected during weight assignment recommended by the used method, Fuzzy Analytic Hierarchy Process (FAHP). Our proposed approach based on FAHP will allow as selecting the most suitable cloud service to users' requirements in terms of big data characteristics while taking into account the dependability

Manuscript received October 26, 2021; revised April 19, 2022.

doi: 10.12720/jait.13.3.203-212

attributes of these data. Then, we implement our approach to prove its efficiency.

The remainder of this paper is organized as follows: in the second section, we will introduce Big Data, the dependability of Big Data, cloud services, and the FAHP method. In Section III, we will present a review of literature about Cloud Service Selection for Big Data's dependability using FAHP. In the fourth one, we will present the model that we propose. Finally, we will finish by concluding and giving some proposals for further research and investigation.

II. Context

A. Big Data

The term Big Data is used to describe data sets that are too large, move too fast, and are complex to work with using commonly available tools. The intrinsic characteristics of Big Data known by 5V are: (i) Data volume refers to the increasing amount or scale of data in terms of size. (ii) Data velocity describes the frequency of data generation, capture, and sharing. (iii) Data have very varied and heterogeneous forms. (iv) Data veracity refers to uncertainty in data, which may or may not be trusted because it may or may not be verified. [7]. (v) Data value represents data utility.

B. Dependability of Big Data

According to Laprie [8], dependability involves different concepts, it is composed of the following three elements: (i) attributes, (ii) constraints, and (iii) means. Attributes are used to specify the expected properties of the system and to assess the quality of the service provided. The means are the methods and techniques to ensure the ability of the system to deliver a service that is consistent with the fulfillment of its function and to provide confidence in that ability. The constraints are related to undesirable but expected circumstances.

(1) Data Availability: This means that users can access the data under given conditions, at a given time.

(2) Data maintainability: is the ability of an asset to be maintained or restored to a condition where it can perform a required function [9].

(3) Data reliability is defined as the probability that an entity will continue to perform a function under given conditions.

(4) Data confidentiality is related to the probability of information leakage from each cloud and the total disclosure level of the information [10].

(5) Data integrity refers to ensuring the authenticity of information accessed: information must not have been altered and must come from a secure source [11].

C. Cloud Services

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [12].

Cloud offers dematerialization services for IT resources such as (i) Software as a Service (SaaS); (ii) Platform as a Service(PaaS); and (iii) Infrastructure as a Service(IaaS) [13].

SaaS: In this type of service, applications are made available to consumers. Applications can be manipulated using a web browser, and the consumer does not have to worry about updating, adding security patches, or ensuring service availability [14].

IaaS: consists of providing access to a virtualized computer park. Virtual machines on which the consumer can install an operating system and applications are made available [15].

PaaS represents an environment that allows the company to deploy its applications outside its computer room. Thus, the company rents a middleware environment with a hidden infrastructure.

D. Fuzzy Analytic Hierarchy Process

AHP method consists of presenting a decision problem in a hierarchical structure and making pairwise comparisons of the elements of the hierarchy [16].

Basic AHP does not include inaccuracy in personal judgments, for which it has been improved using the fuzzy set theory. In FAHP, pairwise comparisons of criteria and alternatives are made using linguistic variables, represented by triangular numbers [17]. Table I shows the Saaty scale.

TABLE I. Saaty Scale

Saaty scale	Definition	Scale
1	Equally important	(1,1,1)
3	Weakly important	(2,3,4)
5	Fairly important	(4,5,6)
7	Strongly important	(6,7,8)
9	Absolutely important	(9,9,9)
2	The intermittent values between two aadjacent scales	(1,2,3)
4		(3,4,5)
6		(5,6,7)
8		(7,8, 9)

Example:

If the decision-maker declares, criterion 1 (C1) is small compared to criterion 2 (C2) then, he takes the fuzzy triangular value of (2, 3, 4). Conversely, the comparison of C2 to C1 takes the fuzzy triangular value (1/4, 1/3, 1/2).

III. Related Works

Several works have discussed cloud service selection methods, while others have focused on Big Data dependability. In this section, we will start by presenting a literature review of the cloud service selection methods, and then we will present works on Big Data dependability, to select works that have proposed solutions for Cloud Service Selection for Big Data's dependability using FAHP.

A. Cloud Service Selection Methods

In the literature, some different methods and techniques deal with the selection of the Cloud service such as a novel fuzzy integrated AHP method based on a hierarchical structure with six principal criteria, and four alternatives [18]. In addition a hybrid multi-criteria decision-making model for a cloud services selection problem using a Balanced Scorecard (BSC), Fuzzy Delphi Method (FDM), and FAHP method [19]. Alam *et al.* developed an integrated decision model based on FAHP and the Weighted Aggregate Sum Product Assessment (WASPAS) [20] to solve the cloud service selection problem. To address this critical selection gap, Cheng presented an integrative search model using FAHP. He used this model to evaluate PaaS and IaaS-based cloud computing providers with predetermined criteria. [21]. Also a new model for selecting cloud services was developed by Kumar *et al.* this model is based on FAHP and Technique for Order Preference by Similarly to Ideal Solution (TOPSIS) [22]. And Kwon & Seo proposed a FAHP-based decision-making model to select an appropriate cloud service provider [23].

B. *Big Data Dependability*

Big Data dependability represents a major interest for the scientific community. Several techniques have been proposed on this subject. Regarding the security of Big data, we find a secure XTS-AES implementation method in the flash translation layer and a SecureXTS proof-of-concept implementation using OpenNFM [24]. A Big Data security fragmentation framework to secure data was proposed by Nabawy *et al.* [25]. To solve the security issues associated with current encryption methods, Reddy proposed a new security scheme called Multilayer Hybrid Encryption Standard [26]. According to [27] the Blockchain may improve Big Data security. In addition, a Lightweight Security Scheme for Big Data based on Elliptic Curve Diffie Hellman was proposed by Sirisha and Kiran [28]. Another approach was proposed in terms of security, which consists of controlling Big [29].

In order to improve data confidentiality and integrity, Bhattacharjee *et al.* developed an integrating approach by reducing data corruption, and processing costs in Big Data transactions [30]. On the other hand, Safhi and al. provide a framework that automatically evaluates the process reliability of Big Data [31]. And Wang and al. deploy a reliability model that enables study the data reliability [32]. An approach was developed to exploit the potential of Big Data and facilitate the construction and maintenance of systems focused on the analysis of large data [33]. To surmount the constraints of the RIF system, and to improve the availability and the reliability; Kaseb *et al.*, proposed a technique called High Availability Redundant Independent Files. Based on a CP, it is called CPHARIF [34]. A new model to analyze the reliability of Big Data was constructed [35]. Wei and al proposed a new generalized signature scheme based on ID using Waters and CHK techniques. The proposed scheme can function as an encryption scheme, a signature scheme, or a signcryption scheme depending on the need [36].

The state of art summarized in Table II highlighted that works studying Big Data dependability have addressed a limited number of attributes. Some work has proposed solutions to improve Big Data security. While others have proposed approaches to develop one of these attributes: reliability, confidentiality, or maintainability. Some researches focus on the combination of two attributes at most.

TABLE II. SUMMARY OF THE LITERATURE REVIEW

Contribution	Cloud Service Selection	Big Data	Dependability	AHP/FAHP
[24]		X	X (Confidentiality)	
[25]	X	X	X (Security)	
[30]		X	X (Confidentiality, Integrity)	
[26]		X	X (Security)	
[28]		X	X (Security)	
[27]		X	X (Security)	
[34]		X	X (availability & Reliability)	
[4]	X			X
[20]	X			X
[19]	X			X
[31]		X	X (Reliability)	
[22]	X			X
[21]	X			X
[35]		X	X (Reliability)	
[36]		X	X (Confidentiality)	
[29]		X	X (Security)	
[23]	X			X

According to the literature review, we found that there is practically non-existent work that deals with the interactions between the 5Vs of Big Data and the dependability attributes (i) Availability, (ii) Confidentiality, (iii) Reliability, (iv) Security, (v) Integrity, (vi) Maintainability. Regarding the Cloud service selection, we have identified problems related to the non-consideration of the specific needs of Big Data.

In addition, the diversity of Cloud services makes the selection very difficult. Therefore, an efficient and precise decision-making approach is highly desirable to select the Cloud service that corresponds to the users' requirements in terms of big data characteristics, while taking into account the dependability.

Our approach aims to integrate the five concepts of dependability (Availability, Confidentiality, Reliability, Security, Integrity, Maintainability.) in the process of cloud services selection adapted to the main characteristics (5V) of Big Data.

IV. CONTRIBUTION

Selecting the appropriate cloud service for Big Data is not the primary challenge for cloud users. Indeed, data dependability is an important aspect. Nowadays, organizations find a major interest in Cloud Services, while the Big Data dependability problem remains a major obstacle. Our main contribution is to integrate the dependability requirements in the selection of the Cloud service adapted to Big Data using FAHP.

FAHP process of choosing CS involves several steps. The most important step is the assignment of the weight.

This, requires a preliminary study of criteria, dependability attributes, and their interactions.

A. *Analysis Study*

In this section, we study different interactions between dependability attributes and the 5V of Big Data to obtain the weight assignment guidelines. These guidelines are an essential element in the process of selecting a cloud that meets the requirements of Big Data dependability.

1) *Interaction between dependability attributes*

As mentioned previously, availability represents the probability that the data will be able to perform their required function when called, and is not subject to repair. Therefore, availability depends not only on reliability but also on ease of maintenance.

Data maintainability is essential to ensure the availability of data for the long term. It ensures data availability despite the occurrence of temporary or permanent unavailability of storage nodes, so that the system may lose some data if no action is taken in case of a failure.

Taking into account both reliability and maintainability, an additional metric is required to determine the probability that the entity will be operational at any given time (i.e., it has not failed or has been restored after a failure); This metric is the availability [9]. Availability is a performance criterion that takes into account reliability and maintainability properties. According to [9], corrective maintenance by reducing the failure rate improves reliability. Fig. 1 shows the components of equipment availability. It highlights:

- That maintainability is one of the levers of action to improve the availability and therefore the productivity of the equipment.

- That reliability and maintainability are 2 parallel notions of equal importance (and whose analysis approaches are similar).

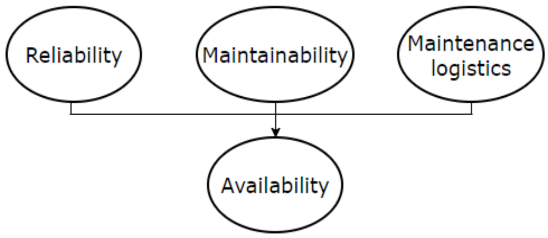

Figure 1. Components of equipment availability.

In this context, Bellaouar made a study to show the effect of maintenance on reliability [9]:

Rm (t) = reliability of the system maintained

t = operating time

T = time at which maintenance is performed

0≤ t ≤T => Rm=R (T), Maintenance does not affect system reliability in [0, T]

t = T, maintenance is performed. The system becomes as good as new. The effect of ideal maintenance on reliability is shown in Fig. 2:

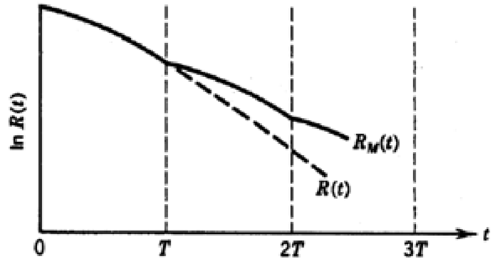

Figure 2. Maintenance effect on reliability.

Note: Reliability implies a notion of duration (operation for a certain duration) while availability implies a notion of immediacy (operation at a given time).

Relationships between reliability, maintainability, and availability are illustrated in Table III [37]. For instance, while maintainability and availability increase, reliability remains constant.

TABLE III. INTERACTIONS BETWEEN DEPENDABILITY ATTRIBUTES

Reliability	Maintainability	Availability
↔	↓	↓
↔	↑	↑
↑	↔	↑
↓	↔	↓

↑: Increase ↔: Constant ↓: Decrease

Note that security is one of the dependability concepts that is considered as a balance between confidentiality, integrity, and availability [38]. The confidentiality problem is easily addressed if availability is not needed. Usually, availability requirements have been ignored when technical solutions to privacy problems have been addressed.

In this way, anonymity could be a solution, because if private data is inaccessible, it will not be misused. For instance, Data confidentiality can be ensured by disconnecting the Database which will make it unavailable [39]. In the e-health context, physicians perceive that a regionally shared drug list increases data availability on current prescriptions but reduces data confidentiality [40].

Even if the notions of integrity and availability are distinct, they are still linked [11]. Indeed, data integrity is only relevant if data is available. Conversely, data availability is relevant only if data are integrated. To preserve data integrity and protect personal data, it must be stored confidentially [41]. Interactions between dependability attributes are summarized in Table IV.

TABLE IV. INTERACTIONS BETWEEN DEPENDABILITY ATTRIBUTES

Confidentiality	Availability	Reliability	Integrity	Maintainability
↑	↓	↑	↑	↔
↓	↑	↔	↓	↑
↑	↑	↑	↑	↔
↔	↔	↑	↑	↔
↔	↑	↔	↔	↑

2) Interaction between 5V of big data

Big Data develops, grows, and changes rapidly. This growth leads to the variable nature of Big Data. As so as, data from multiple sources of different types and representations are highly interrelated [42]. If data variety and data volume increase, the veracity of data decrease. Data veracity refers to the data source.

Value is one of the 5 V's. Big Data storage and analysis technologies make sense only if they provide added value. The other Big Data characteristics are not relevant if these data have no value [43]. The variety of data directly affects their integrity [44]. In other ways, the more the data is varied the more it contains errors. As a result, the veracity of data decreases.

The high processing speed is directly responsible for a high volume of data [15]. All the previous Interactions are summarized in Table V.

TABLE V. INTERACTIONS BETWEEN 5V BIG DATA

Volume	Velocity	Variety	Veracity	Value
↑	↑	↑	↔	↔
↑	↑	↔	↔	↔
↑	↑	↔	↓	↓
↔	↔	↔	↑	↑

3) Interaction between dependability attributes and 5V big data

In the process of selecting a cloud that meets the requirements of Big Data dependability, fixing the guidelines represents a crucial element. Therefore, in this section, we will focus on the interactions between 5V Big data and dependability attributes which represent the weight assignment. It allows the user to assign the appropriate weight to each criterion to choose the most suitable alternative.

The guidelines are illustrated in Table VI. For example, the first row of the table corresponds to the volume: If data volume increases, integrity and maintainability will decrease, while other attributes will not be affected. To maintain a balance between all attributes we must assign an average weight to the volume to not damage the others.

TABLE VI. INTERACTIONS BETWEEN DEPENDABILITY ATTRIBUTES AND 5V BIG DATA

Then / If	Confidentiality	Availability	Reliability	Integrity	Maintainability
Volume ↑	↔	↔	↔	↓	↓
Velocity ↑	↓	↑	↓	↓	↓
Variety ↑	↔	↔	↔	↓	↓
Veracity ↑	↑	↑	↑	↑	↔
Value ↑	↔	↔	↔	↑	↔

B. FAHP Application Process

The proposed solution consists of the selection of cloud services adapted to the main characteristics of Big Data. We will proceed with the following steps:

(i) The first step is to identify the criteria and alternatives. In our approach, we have two categories of criteria: criteria related to Big Data and criteria related to cloud services. Many criteria are involved in cloud service selection. Based on literature review, [10] the most common criteria to choose the right Cloud Service are: performance, scalability, and Cost [45]. The 5V (Volume, Velocity, Variety, Value, and Veracity) represent the Big Data criteria while the cloud criteria are set by the end-user. The alternatives to this problem are Saas, PaaS, and IaaS.

(ii) The second step is to specify the problem hierarchy as shown in Fig. 3.

(iii) In the third step of our approach, we integrate the five attributes of dependability (Availability, Confidentiality, Reliability, Security, Integrity, and Maintainability).

Therefore, we construct a pairwise comparison matrix among decision elements using the FAHP scales presented in Table I, to obtain the average vector. In this step, when assigning weights, we should take into consideration the results obtained in Table VI.

(iv) In step four we fill the pairwise matrix of alternatives for each criterion and calculate the final vector for each matrix.

(v) Finally, we aggregate in the fifth step the relative weights of decision elements to obtain an overall rating of the alternatives, using the vectors obtained from matrices of step four and the average vector obtained from the third step.

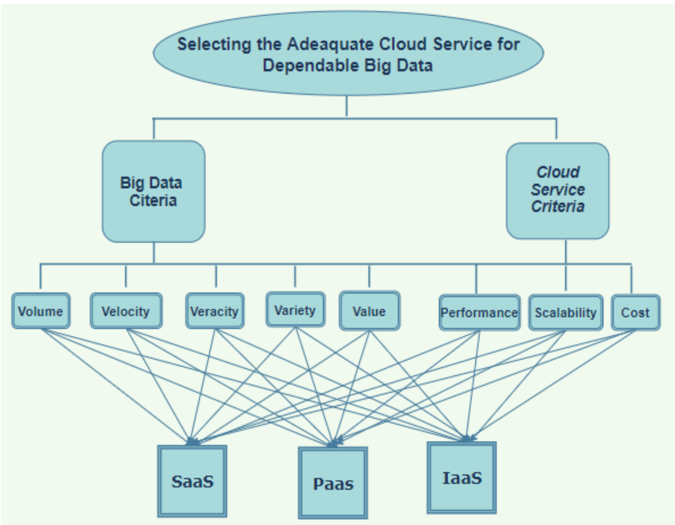

Figure 3. FAHP hierarchy.

C. Simulation and Validation

To evaluate our approach, we conceived and implemented a web platform to automate the selection of the cloud service which respects Big Data dependability requirements. The web platform is developed using XAMPP (X (cross) Apache MariaDB Perl PHP), which allows to set up a local web server apache. Our system is build based on MySQL database, PHP, HTML, CSS, and JavaScript. It can be accessed from any web client anytime the user is connected to the Internet.

To validate its efficacy, we applied our proposed approach to a case study that we have used in previous work [7]. This case study is based on the Big Data characteristics of the National Health Service (NHS) of the United Kingdom (UK), as shown in Table VII [46].

The web platform suggests inserting the cloud criteria. Note that Big Data criteria which are based on NHS Big Data characteristics listed in Table VII are already inserted in our database. And then, it allows us to fill in the alternatives list of cloud services.

TABLE VII. BIG DATA CHARACTERISTICS OF NHS

5Vs of Big Data	NHS's Data Characteristics
Volume	Increasing size
Velocity	Speed of generation and processing
Variety	Heterogeneous Formats (XML, CSV, Multimedia, etc.)
Veracity	Critical authenticity
Value	Significant

After validation of the alternatives list, the comparison matrix will be created. In this step, based on the guidelines of Table VI, the user fills the upper triangular part of the matrix with the weights of criteria of both Big Data and cloud service as mentioned in Fig. 4. And respecting the FAHP scale in Table I; the lower triangular part of the matrix is automatically generated.

The studied case shows that the volume compared to velocity is less important according to NHS requirements; that is the reason behind the fact the fuzzy triangular value of (2, 3, 4) is assigned to it. Consequently, the comparison between velocity and volume takes automatically the fuzzy triangular value (0.25, 0.33, 0.5). (see Fig. 4).

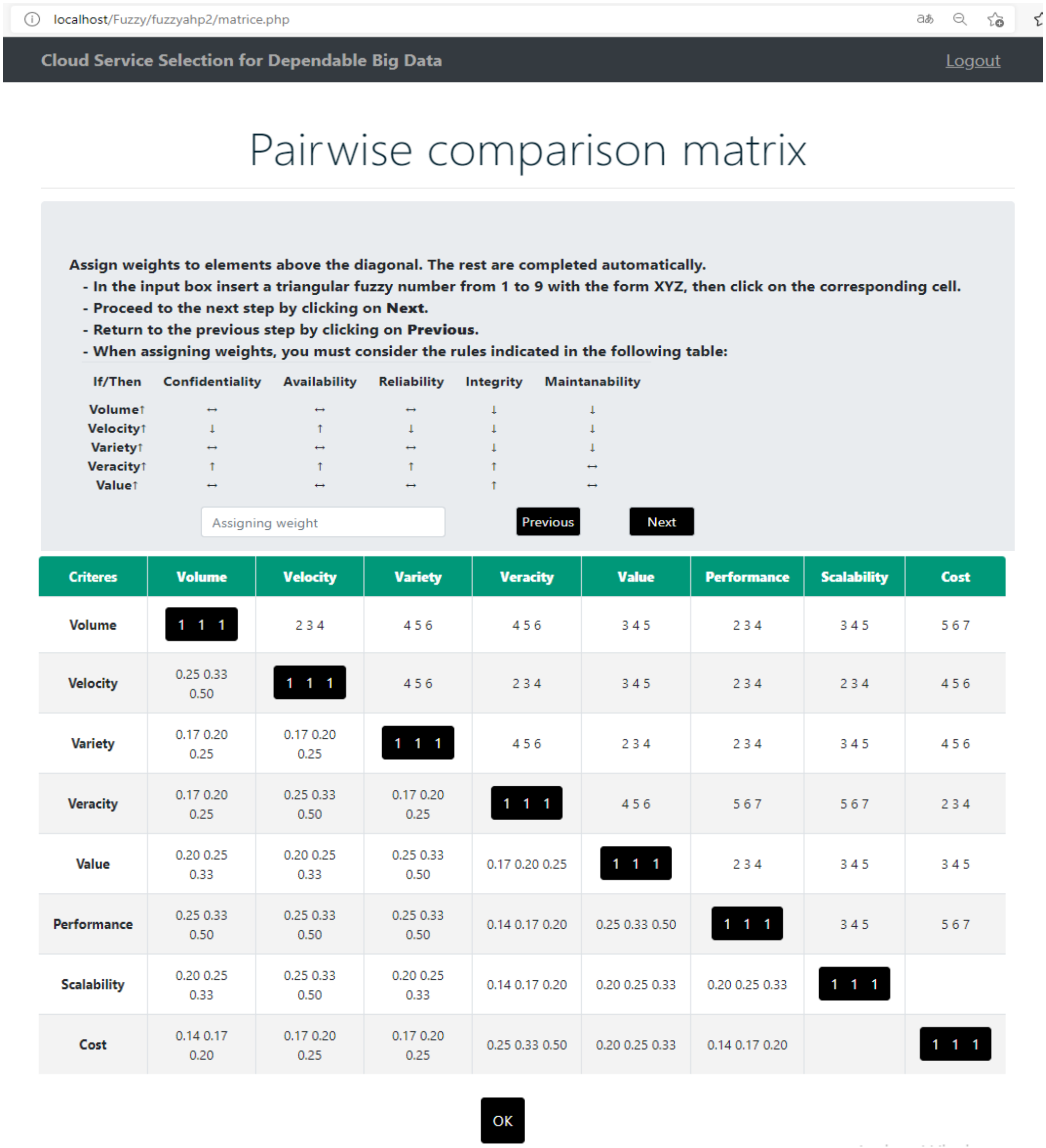

Figure 4. Pairwise comparison matrix.

After validation of the matrix, various mathematical calculations (Fig. 5, Fig. 6, and Fig. 7) are performed to obtain the average vector (F3 = Average (F2)) calculated using (i) F1: function to calculate the percentage of each criterion and (ii) F2= ∑ F1. Fig. 5 represents the function applied to calculate the percentage of each pairwise comparison criteria (F1). Fig. 6 shows the function applied to calculate the sum of the percentage of each matrix line (F2). Fig. 7 shows the function used to calculate the average of sum of the percentage of each matrix line (F3).

```
function calculeMatrice(elm){
    id = elm.id
  var var1 = '0 0 0'
    for(var i = 0;i < limit; i++){
      var1 = '0 0 0'
      for(var j=0;j < limit; j++){
        var tabTemp = matrice[id][j][i].split(' ')
        var valeurTemp = var1.split(' ')
        var var2 = Number(tabTemp[0]) + Number(valeurTemp[0])
        var var3 = Number(tabTemp[1]) + Number(valeurTemp[1])
        var var4 = Number(tabTemp[2]) + Number(valeurTemp[2])
        var1 = var2 + ' ' + var3 + ' ' + var4  }
      sum[id].push(var1)    }
    console.log(sum)
    return pourcentageCri(id) }
```

Figure 5. Function to calculate the criteria percentages.

```
function pourcentageCri(id){

    for(var i = 0; i< limit ;i++) {
      matriceCri [id][i] = [];

        for(var j = 0; j< limit ;j++) {
          var cri1 = (matrice[id][i][j]).split(' ')
          var cri2 = (sum[id][j]).split(' ')
          var temp1 = Number(cri1[0])/Number(cri2[0])
          var temp2 = Number(cri1[1])/Number(cri2[1])
          var temp3 = Number(cri1[2])/Number(cri2[2])

          var varTemp = temp1 + ' ' + temp2 + ' ' + temp3

          matriceCri[id][i][j]= varTemp

        }
    }

    console.log(matriceCri)
    return sumCri(id);
  }
```

Figure 6. Function to calculate the criteria percentage sum.

```
function sumCri(id){

  for(var i = 0; i< limit ;i++) {
    var init = "0 0 0"
      for(var j = 0; j< limit ;j++) {
          var temp  =  (matriceCri[id][i][j]).split(' ')
          var tempInit = init.split(' ')

          var temp1 = Number(temp[0]) + Number(tempInit[0])
          var temp2 = Number(temp[1]) + Number(tempInit[1])
          var temp3 = Number(temp[2]) + Number(tempInit[2])

          init = temp1 +' '+ temp2 +' '+ temp3

      }
      tabSumCri[id].push(init)
  }
  console.log(tabSumCri)
  return sumCriN(id);
}
```

Figure 7. Function to calculate the criteria average.

In this step, the user is invited to compare the alternatives for each criterion by assigning weight for alternatives. These weights are inserted in the upper triangular part of the matrices. The lower triangular part of the matrices is automatically generated.

After validation of all matrices, various mathematical calculations are performed to obtain the vectors of alternatives. The final result is based on the average vector (the result of F3), and the vectors of alternatives as shown in Fig. 8.

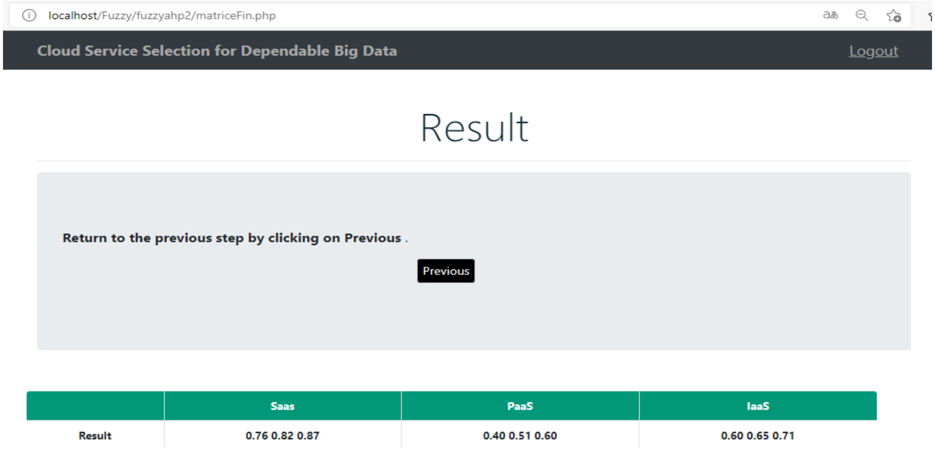

Figure 8. Interface for adequate cloud service.

The final result shows that in the case of the NHS, the most suitable cloud service for Big Data, respecting the dependability requirements is SaaS.

Thus, our contribution allows us to get the best choice when selecting cloud services adequate to dependable Big Data.

D. *Evaluation of Our Approach*

To evaluate the ability of our platform to perform a large number of criteria, we have done some simulations. The evaluation of the proposed approach is performed by varying the number of criteria from 6 to 50 for 3 alternatives and 20 alternatives. Then, we measured the execution time, and we obtained the results shown in Table VIII and Fig. 9. The simulations were performed on a host with a windows 10 system (64 bit) with an i3 CPU 2.60GHz, and a 4GB RAM.

These results demonstrated that if we vary the number of criteria the execution time remains unchanged. These results prove that our platform is independent of the number of criteria. The user can input all the criteria needed to choose a suitable cloud service with a reasonable execution time.

TABLE VIII. EXECUTION TIME/ NUMBER OF CRITERIA

Number of criteria	Execution time 3	execution time 20
6	8	10
15	19	20
20	22	24
25	26	27
30	29	32
35	33	37
40	39	41
45	44	45
50	49	50

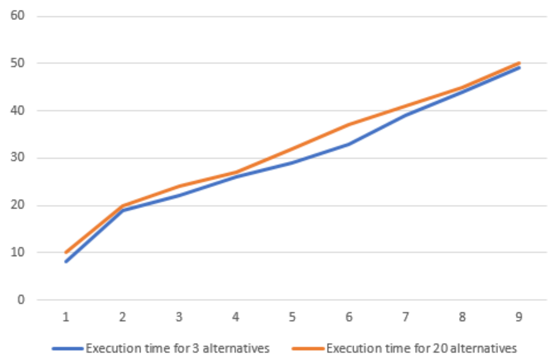

Figure 9. Execution time/Number of criteria.

V. CONCLUSION AND PERSPECTIVES

In this article, we propose an approach that aims to integrate dependability concepts into the process of selecting the cloud service adapted to Big Data. Our contribution is based on the analysis study of dependability attributes, the 5V of Big Data. We study different interactions between dependability attributes and the 5V of Big Data in order to obtain the weight assignment guidelines that will be used in FAHP.

In order to validate our approach, we implement a platform, which proves its feasibility. Then we evaluate the implemented platform using the NHS case study. As result, we obtain the best cloud alternative that meets the dependability requirement of NHS Big Data.

In future work, we intend to expand the list of Big Data characteristics considered. We aim to provide a set of weighting guidelines that encompass other characteristics of Big Data to keep tracking the evolution of this field. We also plan to make our platform available to the cloud providers, which will allow them to automatically put on their various services. Therefore, the implemented system will become an exchange platform between users and cloud providers.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Fatima Ezzahra Mdarbi conducted the research, analyzed the data and wrote the paper. All of the authors have contributed to validate the results obtained and approve the final version of this paper.

REFERENCES

[1] S. Yin and O. Kaynak, "Big data for modern industry: Challenges and trends [point of view]," *Proceedings of the IEEE*, vol. 103, no. 2, pp. 143-146, 2015.

[2] X. Li, Y. Zhuang, and S. X. Yang, "Cloud computing for big data processing," *Intelligent Automation & Soft Computing*, vol. 23, no. 4, pp. 545-546, Oct. 2017.

[3] F. E. Mdarbi, N. Afifi, and I. Hilal, "Big data dependability opportunities challenges," in *Proc. 1st International Conference on Smart Systems and Data Science*, Oct. 2019, pp. 1-4.

[4] R. R. Kumar and C. Kumar, "An evaluation system for cloud service selection using fuzzy AHP," in *Proc. 11th International Conference on Industrial and Information Systems*, 2016, pp. 821-826.

[5] N. Uraipan, P. Praneetpolgrang, and T. Manisri, "Application of an analytic hierarchy process to select the level of a cyber resilient capability maturity model in digital supply chain systems," *ECTI Transactions on Computer and Information Technology*, vol. 15, no. 2, 2021.

[6] K. Bennis and L. Bahi, "Application of fuzzy logic to participatory decision support for environmental impact assessment – Case of Tangier," *Journal of Decision Systems*, vol. 25, no. 1, pp. 56-78, 2016.

[7] F. E. Mdarbi, N. Afifi, I. Hilal, and H. Belhadaoui, "An approach for selecting cloud service adequate to big data case study: E-health context," *International Journal of Computer Science and Information Security*, vol. 16, no. 8, August 2018.

[8] J. C. Laprie, "Dependable computing: Concepts, limits, challenges," in *Proc. the 25th IEEE International Symposium on Fault-Tolerant Computing*, Pasadena, California, USA, 1995.

[9] *Fiabilite Maintenabilite Disponibilite*, P. A. Bellaouar, 2013, p. 106.

[10] Y. Kajiura, S. Ueno, A. Kanai, S. Tanimoto, and H. Sato, "An approach to selecting cloud services for data storage in heterogeneous-multicloud environment with high availability and confidentiality," in *Proc. IEEE Twelfth International Symposium on Autonomous Decentralized Systems*, 2015, pp. 205-210.

[11] ASIP, *Guide to Mechanisms for Protecting the Integrity of Stored Data*, The French Digital Health Agency, 2017.

[12] T. Grance and P. Mell, "The NIST definition of cloud computing," Special Publication 800-145, Oct. 25, 2011.

[13] S. Lee, H. Park, and Y. Shin, "Cloud computing availability: Multi-clouds for big data service," in *Proc. International Conference on Hybrid Information Technology*, 2012, pp. 799-806.

[14] Memory Online - Study on the security of cloud computing - Moussa MIMOUNE. [Online]. Available: https://www.memoireonline.com/07/15/9194/Etude-sur-la-securite-du-cloud-computing.html

[15] Q. Machu, "A datacenter in the public cloud/HA IaaS," Polytechnic School of François Rabelais University of Tours, 2015.

[16] T. L. Saaty, "Decision making with the analytic hierarchy process," *International Journal of Services Sciences*, vol. 1, no. 1, p-. 83-98, 2008.

[17] T. L. Saaty, "Group decision making and the AHP," in *The Analytic Hierarchy Process*, Berlin, Heidelberg: Springer, 1989, pp. 59-67.

[18] I. Otay and T. Yıldız, "Multi-criteria cloud computing service provider selection employing pythagorean fuzzy AHP and VIKOR," in *Proc. International Conference on Intelligent and Fuzzy Systems*, 2020, pp. 423-431.

[19] S. Lee and K. K. Seo, "A hybrid multi-criteria decision-making model for a cloud service selection problem using BSC, fuzzy Delphi method and fuzzy AHP," *Wireless Personal Communications*, vol. 86, no. 1, p. 57-75, 2016.

[20] K. A. Alam, R. Ahmed, F. S. Butt, S. G. Kim, and K. M. Ko, "An uncertainty-aware integrated fuzzy AHP-WASPAS model to evaluate public cloud computing services," *Procedia Computer Science*, vol. 130, pp. 504-509, 2018.

[21] X. Cheng, "Cloud computing decision-making using a fuzzy AHP approach," *Management Information Systems*, vol. 20, no. 4, pp. 89-116, 2015.

[22] R. R. Kumar, S. Mishra, and C. Kumar, "Prioritizing the solution of cloud service selection using integrated MCDM methods under Fuzzy environment," *The Journal of Supercomputing*, vol. 73, no. 11, pp. 4652-4682, 2017.

[23] H. K. Kwon and K. K. Seo, "A decision-making model to choose a cloud service using fuzzy AHP," *Advanced Science and Technology Letters*, vol. 35, pp. 93-96, 2013.

[24] Q. Zhang, *et al*., "Ensuring data confidentiality with a secure XTS-AES design in flash translation layer," in *Proc. IEEE 5th International Conference on Cloud Computing and Big Data Analytics*, 2020, pp. 289-294.

[25] R. M. Nabawy, H. E. Beh, and H. M. Mousa, "Securing big data using mixed fragmentation based on multi-cloud environment," *International Journal of Scientific & Technology Research*, vol. 9, no. 3, p. 9, 2020.

[26] P. A. Reddy, P. B. Reddy, and O. Ramesh, "A novel scheme for security of unstructured and semi-structured big data outsourced to cloud," *Computer Science*, p. 6, 2020.

[27] S. Wibowo and A. D. W. Sumari, "The utilization of blockchain for enhancing big data security and veracity," in *Combating Security Challenges in the Age of Big Data: Powered by State-of-the-Art Artificial Intelligence Techniques*, Z. Md. Fadlullah and A. S. Khan Pathan, Eds., Cham: Springer International Publishing, 2020, pp. 157-187.

[28] N. Sirisha and K. V. D. Kiran, "An efficient and lightweight security scheme for big data," *International Journal on Emerging Technologies*, vol. 11, p. 7, 2020.

[29] C. Tankard, "Big data security," *Network Security*, vol. 2012, no. 7, pp. 5-8, 2012.

[30] S. Bhattacharjee, L. B. A. Rahim, J. Watada, and A. Roy, "Unified GPU technique to boost confidentiality, integrity and trim data loss in big data transmission," *IEEE Access*, vol. 8, pp. 45477-45495, 2020.

[31] H. M. Safhi, B. Frikh, and B. Ouhbi, "Assessing reliability of big data knowledge discovery process," *Procedia Computer Science*, vol. 148, pp. 30-36, 2019.

[32] Z. Wang, C. Cao, N. Yang, and V. Chang, "ABE with improved auxiliary input for big data security," *Journal of Computer and System Sciences*, vol. 89, pp. 41-50, 2017.

[33] A. Kumar, F. Niu, and C. Ré, "Hazy: Making it easier to build and maintain big-data analytics," *Queue*, vol. 11, no. 1, pp. 30-46, 2013.

[34] M. R. Kaseb, M. H. khafagy, I. Ali, and S. Essayed, "An improved technique for increasing availability in big data replication," *Future Generation Computer Systems*, vol. 91, pp. 493-505, 2019.

[35] X. Wu, X. Liu, and S. Dai, "The reliability of big data," in *Proc. IEEE 7th Joint International Information Technology and Artificial Intelligence Conference*, 2014, pp. 295-299.

[36] G. Wei, J. Shao, Y. Xiang, P. Zhu, and R. Lu, "Obtain confidentiality or/and authenticity in big data by ID-based generalized signcryption," *Information Sciences*, vol. 318, pp. 111-122, 2015.

[37] Reliability HotWire. (2003). Relationship between availability and reliability. [Online]. Available: https://www.weibull.com/hotwire/issue26/relbasics26.htm

[38] I. A. Sumra, H. B. Hasbullah, and J. B. AbManan, "Attacks on security goals (confidentiality, integrity, availability) in VANET: A survey," in *Vehicular Ad-Hoc Networks for Smart Cities*, Singapore: Springer, 2015, pp. 51-61.

[39] M. S. Olivier, "Database privacy: Balancing confidentiality, integrity and availability," *ACM SIGKDD Explorations Newsletter*, vol. 4, no 2, pp. 20-27, 2002.

[40] T. Hammar, A. Ekedahl, and G. Petersson, "Implementation of a shared medication list: Physicians' views on availability, accuracy and confidentiality," *International Journal of Clinical Pharmacy*, vol. 36, no. 5, pp. 933-942, 2014.

[41] T. R. Fleming, K. Sharples, J. McCall, A. Moore, A. Rodgers, and R. Stewart, "Maintaining confidentiality of interim data to enhance trial integrity and credibility," *Clinical Trials*, vol. 5, no. 2, p.. 157-167, 2008.

[42] N. Zanoon, A. Al-Haj, and S. M. Khwaldeh, "Cloud computing and big data is there a relation between the two: A study," *International Journal of Applied Engineering Research*, vol. 12, no. 17, pp. 6970-6982, 2017.

[43] I. Hausmane. (2019). Les 10 "V" du big data - Analytics & insights - Big data", analytics & insights. [Online]. Available: https://analyticsinsights.io/les-10-v-du-big-data/

[44] A. M. Khan, Ü. C. Büyüksahin, and F. Freitag, "Prototyping incentive-based resource assignment for clouds in community networks," in *Proc. IEEE 28th International Conference on Advanced Information Networking and Applications*, 2014, pp. 719-726.

[45] R. K. Tiwari and R. Kumar, "G-TOPSIS: A cloud service selection framework using Gaussian TOPSIS for rank reversal problem," *The Journal of Supercomputing*, vol. 77, no. 1, pp. 523-562, 2021.

[46] (2017). Can big data analytics cure the NHS? [Online]. Available: https://www.capgemini.com/gb-en/2017/06/can-big-data-analytics-cure-the-nhs/

Fatima Ezzahra Mdarbi is a PhD student at Networks, Computer, Telecom & Multimedia (RITM) Laboratory at the Higher School of Technology at Hassan II University of Casablanca Morocco (UH2C). Doctoral study center engineering science of the National Higher School of Electricity and Mechanics (ENSEM) at UH2C. She is graduate as a computer engineer from the National School of Applied Sciences of Khouribga in 2015. Her research interests include Big Data, cloud computing and dependability. She participated in many conferences such as BDCA17, SysCoBIoTS19, ICSSD19, and COC19.

Nadia Afifi is a Professor at Higher School of Technology (ESTC), Hassan II University of Casablanca Morocco (UH2C) since 1993. She received an Electrical Engineering, Computer Systems degree from the Mohammadia School of Engineers (EMI), Mohammed V University - Rabat (1988). She received a PhD in Computer Engineering from UH2C; and HDR in IT in 2014. She is a Senior Member; part of "Software Engineering, Business Intelligence and Security" (SEBIS) research team of "Networks, Computer, Telecom & Multimedia" (RITM) Laboratory (ESTC). She is member of the "Moroccan Association of Information Technology" (AMTI), and member of UH2C Chair on Innovation (UNCHAIN). Her research interests include: Software Architecture; Information Systems Security, Information Systems Modeling, Human Machine Interface, Dependability, Business Intelligence; Cloud Computing Security, Big Data, IoT and Artificial Intelligence. She participated as a consultant then software architect in several national and international projects. She developed several partnerships with international universities, notably in Korea Republic. She coordinated and participated in several international cooperation programs.

Imane Hilal is a Professor in the School of Information Sciences (ESI) and cosupervises several theses. She holds a Dipl.-Ing. (MSc) from National School of Computer Science and Systems (ENSIAS) in 2009. Then she spent more than eight years in teaching IT development to IT Specialized technician. In 2016 she got a Ph.D. degree from Hassan II University of Casablanca. Her research interests lie in various Computer Science Areas such as: Data Warehousing, Dependability, MDE, Elearning, Big Data, Cloud Computing. She participated in many conferences as Organising Committee or TPC such as: ISI17, MISC18, ArabWIC19 and as Program Committees Chairs in ICSSD19.