Evaluating Cloud Services using Multicriteria Decision Analysis

Pedro Costa

[1] Department of Computer Science and Engineering, Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal
pedro.c.costa@ist.utl.pt

Abstract. The potential of Cloud services for cost reduction and other benefits has been capturing the attention of organizations. However, a difficult decision arises when an IT manager has to select a Cloud service provider because there are no established guidelines to help make that decision. When several, and often conflictive, criteria should be taken into account to compare different Cloud services from multiple providers, the decision is even more difficult. In order to address this problem, we propose a multi-criteria value model to evaluate Cloud services using a multi-criteria decision analysis (MCDA) approach based on the MACBETH method. The proposed method was demonstrated in a City Council in Portugal to evaluate and then compare two Cloud services: Google Apps and Microsoft Office 365. The proposed method was evaluated with Moody and Shanks Quality Framework, and the evaluation shows the proposed method is suitable for evaluating Cloud services.

Keywords: Cloud Services, IT Services, Multiple Criteria Decision Analysis, MACBETH, and Service Quality.

1 Introduction

Outsourcing is an almost routine part of management and represents in many major organizations and government agencies the greater percentage of their Information Technology (IT) expenditure [1]. However, the IT industry is evolving and there is a new business model which is revolutionizing and changing this industry: Cloud services. Organizations can now contract their services from the Cloud rather than owning the assets to provide those services [2].

Notwithstanding, despite the growing maturity and adoption of Cloud services, most decision makers continue to express some concerns [3]. These concerns show they have doubts about what, when, and how they should migrate to the Cloud because there are no guidelines in this area [4].

Therefore, the problem we address in this paper is that a decision maker (DM) has difficulties to evaluate Cloud services in order to purchase them, thus usually postponing the decision to adopt Cloud services.

In order to address this problem we propose a Multiple Criteria Decision Analysis (MCDA) approach [5], based on MACBETH method [6], [7], [8], to build a multi-criteria value model [9], [10] to evaluate Cloud services.

This paper describes the building process of the proposed multi-criteria evaluation model that was demonstrated in a Portuguese City Council that wanted to migrate their productivity software (mail, office) to the Cloud but did not know how to evaluate the options. The Cloud services evaluated and compared were Google Apps and Microsoft Office 365. At the end of the process we obtained an overall value score for each option, which depicted their overall attractiveness for the City Council.

In order to evaluate our proposal, we used the feedback of the DM during the process and the Moody and Shanks Framework [11]. After the evaluation we concluded the proposed method is suitable for evaluating Cloud services.

This study was conducted by using Design Science Research Methodology (DSRM) that aims at creating a commonly accepted framework for research in Information Systems (IS) as well as creating and evaluating artefacts to solve relevant organization problems [12]. The steps of DSRM that are used to organize the paper are: problem identification and motivation; objectives of a solution definition; design and development; demonstration; evaluation; and communication [13].

2 Problem

This section presents the problem identification and motivation step of DSRM, which defines a specific research problem and justifies the value of the solution.

Cloud Computing is experiencing a strong adoption in the market and this trend is expected to continue [14], but it is still in its beginning and quite far from mature. Therefore, it remains unclear whether fears of Cloud are reasonable in the long term or not and if total cost of ownership (TCO) is favourable to customers when compared to owning systems. Hereupon, risks and costs integration with the Cloud is one of the top concerns that DMs in most organizations have about Cloud. The stigma on Cloud is related to key challenges like security and privacy, offshore data housing, lock-in, and compliance [1].

However, when assessing the challenges of Cloud, there is a need to avoid focusing wholly on these concerns, because the technology is changing, legislation is uncertain, and decisions must be taking into account multiple criteria to address the business and strategic objectives of organizations. In addition, DMs weighed risks differently and this points to a major challenge in organization as they move to Cloud: How to bridge the Cloud 'risk perception gap' that exists in the organizations between business and IT executives [1]. Furthermore, most organizations do not talk on the record, and some information about Cloud mistakes is relayed second hand by the Cloud vendors themselves [15].

Our research problem is about the doubts that organizations have about Cloud adoption and what way they should follow now since they do not have knowledge about the real benefits, risks, and costs. Furthermore, the resistance to change and the unknown of Cloud solutions are connoted as more investment and risk. As a result, DMs are postponing the decision to migrate their services to Cloud. However, the organizations need a systematic solution to evaluate and review their business needs and weigh the potential gains and opportunities by the Cloud against the challenges and risks, to make a well-planned and understood strategy [4].

Our main motivation behind this research problem is that a well-founded proposal may have a big impact in the IT service market concerning benefits, costs and value creation. Despite being a highly discussed topic, there is not a consensual opinion and criteria to address this problem.

In summary, in this research we are interested in developing a method that may be used to help a DM to evaluate Cloud services.

3 Related Work

This section covers the definition of the objectives of a solution of DSRM, in which we infer the goals of the solution from the problem definition and related work. We are going to start with an overview about Cloud Computing, which has different characteristics that must be taken into account (Section 3.1). Then, we are going to give an overview of the existing evaluation models with we can compare the Cloud services (Section 3.2). Afterwards, we are going to give a brief description about MCDA approaches (Section 3.3). Finally, we are going to explain the objectives of the solution (Section 3.4).

3.1 Cloud Computing

The emergence of the Cloud represents a fundamental change in the way IT services are designed, developed, deployed, scaled, updated, maintained and paid for. The promise of Cloud is to deliver the functionality of existing IT services even as it dramatically reduces the upfront costs of computing that deter many organizations from deploying many cutting-edge IT services [16].

Cloud Computing is defined by the National Institute of Standards and Technology (NIST) [17], and is mostly used by researchers. Also to NIST, there are four deployment models: private Cloud; community Cloud; public Cloud; and hybrid Cloud. This deployment models are offered in three service models: Software as a Service (SaaS); Platform as a Service (PaaS); and Infrastructure as a Service (IaaS). Finally, for NIST, Cloud has five essential characteristics: on-demand self-service; broad network access; resource pooling; rapid elasticity; and measured service [17].

Traditionally, small and medium enterprises (SME) have to make high capital investment upfront for procuring IT infrastructure, skilled developers and system administrators, which results in a high TCO [18]. Based on this, and knowing that the economic appeal of Cloud is often described as converting capital expenses to operating expenses (CapEx to OpEx), the phrase "pay-as-you-go" can more directly captures the economic benefit to the buyer [4].

DMs have four main desires when they are looking for Cloud adoption. These desires are on a 'desired framework' [1], [19], and are related to: equivalence; abstraction; automation; and tailoring. This framework strips out the value-added benefits of Cloud allowing organizations to focus on the specific differences, and thus make decisions alongside the promised benefits.

Nevertheless, Cloud has four challenges that organizations have to examine deeply. Those challenges are: the weighing up of he security and legal risks; the definition of the contract; the lock-in dilemma; and managing the Cloud [1], [19].

3.2 Evaluation Models

In this subsection we are going to summarize the main existing methodologies and models that try to make the assessment or propose criteria to evaluate Cloud services.

Information and Communication Technology Service Quality (ICTSQ). ICTSQ is a model that has been developed concerning the quality of service (QoS) and is focused on performance measurement. This model only has components related to operational and control performance. While the operational performance is defined as an accomplishment of routine functioning and activities of an organization, the control performance is an accomplishment of the control process to ensure the functions, such as plan, coordinate and monitor, are orderly implemented. These two components still have: variable, which describes the component; factor that is defined as the measurement for variable to indicate the achievement from the evaluation of performance; and attribute, which determines metrics to this evaluation, and the measures that guide the performance score to be given [20].

ISO/IEC 9126 – Information technology – Software product evaluation – Quality characteristics and guidelines for their use. This International Standard defines six characteristics that describe software quality and provide a baseline for further refinement as well as the description of software quality and guidelines that describe the use of quality characteristics for the evaluation of software quality. These characteristics are: *(i)* functionality, which bears on the existence of a set of functions and their specified properties; *(ii)* reliability, which bears on the capability of software to maintain its level of performance under stated conditions for a stated period of time; *(iii)* usability, which bears on the effort needed to use; *(iv)* efficiency, which bears on the relationship between the level of performance of the software and the amount of resources used; *(v)* maintainability, which bears on the effort needed to make specified modifications; and *(vi)* portability, which bears on the ability of software to be transformed from one environment to another [21].

Application Performance Index (APDEX). APDEX is an open standard that defines a standardized method to report, benchmark, and track application performance. It aims to help IT organizations to measure the service levels and customer satisfaction with insight into how well their applications perform from a business point-of-view based on response time, averaging samples, time values, and other measures. For this, APDEX proposes a numeral measure of user satisfaction with the performance enterprise applications. It converts many measurements into one number on a uniform scale from 0 to 1. APDEX is a first user experience metric that alone one to compare performance across application or other reporting groups defined. This metric can be easily understood and used to manage IT across many applications. However, the greatest benefit is its ability to quickly show the alignment of application performance to the business needs [22], [23].

eSourcing Capability Model – Client Organization (eSCM-CL). eSCM-CL has the best practices model, which allows client organizations to continuously evolve, improve, and innovate their capabilities to develop better relationships with their services providers, and to meet the dynamic demands of their business while managing the providers' delivery. This capability model has two purposes: give client organizations guidance that will help them improve their sourcing capabilities throughout the sourcing life-cycle; and provide client organizations with an objective means of evaluating their capability. Additionally, eSCM-CL gives client organizations a standard to use in these evaluations to achieve certification [24].

Service Measurement Index (SMI). SMI is a set of business-relevant KPIs that provide a standardized method for measuring and comparing a business service regardless of whether that service is internally provided or sourced from an outside company. It is designed to become a standard method to help organizations measure Cloud-based business services grounded on their specific business and technology requirements. SMI is a hierarchical framework, the top level divides the measurement space into seven categories that answer to different questions, and each category is further refined by three or more attributes [25]. These categories are [26]: *(i)* accountability: can we count on the provider organization?; *(ii)* agility: can it be changed and how quickly can it be changed?; *(iii)* assurance: how likely is it that the service will work as expected?; *(iv)* financial: how much is it?; *(v)* performance: does it do what we need?; *(vi)* security and privacy: is the service safe and privacy protected?; and *(vii)* usability: is it easy to learn and to use?.

Service Measurement Index Cloud Framework (SMICloud). Based on SMI, SMICloud offers a comparative evaluation framework of Cloud services and helps Cloud customers find the most suitable Cloud provider. It provides features such as service selection based on quality of service (QoS) requirements and ranking of services based on previous user experiences and services' performance. It is a decision making tool, designed to

provide assessment of Cloud services in terms of key performance indicators (KPI) and user requirements. To develop this model they tracked the challenges: how to measure several SMI attributes of a Cloud services; and how to rank the Cloud services based on attributes [18].

Regardless of the choice of sourcing kind, all organizations have to measure and/or evaluate the services. According to several factors the methodologies and frameworks referenced above, and other literature contents less relevant, do not resolve our research problem. The main flaws are represented in **Table 1**.

Table 1. Mapping the flaws with the actual models

Flaws	ICTSQ	ISO/IEC 9126	APDEX	eSCM-CL	SMI	SMICloud
No criteria				✗		
Missing or Redundant Criteria	✗	✗			✗	✗
No Comparative Importance	✗	✗			✗	
KPIs with Unclear Formulas	✗		✗			✗
Complexity			✗	✗		

3.3 Multiple Criteria Decision Analysis (MCDA)

A decision problem typically involves balancing multiple, and often conflicting, criteria. Multiple criteria decision analysis (MCDA) consists in "a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter" [5]. In this subsection we are going to explain briefly some of the most used MCDA methods.

Outranking Methods. Outranking methods are applied directly to partial preference functions, which are defined for each criterion. These preference functions may corresponds to natural attributes on a cardinal scale, or may be constructed in some way, as ordinal scales, and do not need to satisfy all of the properties of value functions, only the ordinal preferential independence would still necessary. In outranking methods, for two alternatives a and b, where $z_i(a) \geq z_i(b)$ for all criteria i, we can say that a outranks alternative b if there is sufficient evidence to justify a conclusion that a is least as good as b, taking all criteria into account [5].

Analytical Hierarchy Process (AHP). AHP is a method based on evaluating alternatives in terms of an additive preference function. The initial steps in using the AHP are to develop a hierarchy of criteria (value tree) and to identify alternatives. AHP uses pairwise comparisons of alternatives to score the alternatives on each criterion and uses pairwise comparison of criteria to weight the criteria, assuming ratio scales for all judgments. The overall score of an alternative is obtained by the weighted summation of its scores on the different criteria [5], [27].

MACBETH. MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) is an approach for multicriteria value measurement [9], [10]. It uses semantic judgments about the differences in attractiveness of several stimuli to help a DM quantify the relative attractiveness of each alternative. It employs an initial, iterative, questioning procedure that compares two elements at a time, requesting only a qualitative preference judgment. As the answers are entered into the MACBETH decision support system [28] it automatically verifies their consistency. It subsequently generates a numerical scale by solving a linear programming problem that is representative of the DM's judgments. Through a similar process it permits the generation of weighting scales for criteria [6], [7], [8].

Outranking methods differ from the others in that there is no underlying aggregative value function, so they do not produce an overall preference scale for each of the alternatives.

AHP generates global scores to represent the overall preference upon the alternatives, which is a wanted feature. However, there are known issues regarding this method concerning, for example, the appropriateness of the conversion from the semantic to the numeric scale used in AHP [29], [30].

A MACBETH advantage over other methods for multi-criteria value measurement is that it only requires qualitative judgments to score alternatives and to weight criteria, instead of requiring quantitative judgments. Furthermore, its decision support system (*www.m-macbeth.com*) is able to compute the overall values of the alternatives by applying the additive model, and to make extensive sensitivity and robustness analysis.

3.4 Objectives

Looking back to the problem we stated (see Section 2), which the main issue is that there is not a consensual set of criteria to evaluate and compare Cloud services. Following the line of thought also concluded that there are no guidelines and structured form to make a complete assessment to Cloud services. Hereupon, the main objective of this proposal is present some mechanism that enables any organization to address this problem. The secondary objectives we intent to achieve with our proposal are: *(i)* clarify DMs doubts and fears about Cloud Computing; *(ii)* easy to apply and not requiring specialized expertise; *(iii)* able to provide understandable results; *(iv)* more affordable than current methods; and *(v)* less expensive than current solutions.

4 Proposal

This section corresponds to the design and develop phase of DSRM, in which we create the artefact. This artefact aims at solving the stated problem.

To address the problem stated (see Section 2), multiple independent criteria must be taken into account to evaluate the Cloud services. In our proposal, we use the MACBETH method to evaluate the options against the criteria previously approved by the DM.

Our method consists in three main steps summarized below:

A) Structuring the model. The decision-making process begins by structuring the problem, which consists in identifying the issues of concern for the DM. The DM fundamental points of view should be taken as evaluation criteria. Each criterion should be associated with a (qualitative or quantitative) descriptor of performance, to measure the extent to which the criterion can be satisfied. We created a template with the descriptors of performance for all Cloud services evaluation criterion presented in [31] (see Table 2). In any case, a DM may always select other evaluation criteria or descriptors of performance in order to meet specific organization's needs.

B) Evaluating the alternatives. In this second step the DM is asked about their preferences in order to build a value function for each criterion and to weight the criteria. The starting point is the definition of two reference levels on each descriptor of performance ("neutral" and "good"). Afterwards, the DM is asked to judge the differences in attractiveness between each two levels of performance by choosing one (or more) of the MACBETH semantic categories: very weak, weak, moderate, strong, very strong, or extreme. Then, M-MACBETH uses a linear programming problem to generate a numerical value scale compatible with the DM's judgments.

To weight the criteria, the DM ranks the neutral–good swings of the criteria by their overall attractiveness. Afterwards, the DM is asked to judge the difference in attractiveness between each two neutral–good swings using the MACBETH semantic categories, and his answers are used by M-MACBETH to create a weighting scale. Finally, the DM should validate the proposed weights and adjust them if necessary.

C) Analysing the results. In this step the performances of the alternatives (factual data) are converted into value scores, using the value functions previously built for each criterion, and an overall value score is calculated for each alternative by weighted summation of its value scores. A final ranking of the alternatives is then achieved using their overall scores. Before giving a selection recommendation it is wise to perform sensitivity and robustness analyses, to know how sensitive or robust is the ranking obtained to "small" changes in the parameters of the model.

The application of the proposal is shown in the next section, where we are going to present a step-wise view of the method in real case.

5 Demonstration

This section corresponds to the demonstration step of DSRM, in which we demonstrate that the proposal can be used to solve one or more instances of the problem.

The main objective of this proposal is constructing some mechanism that enables any organization the evaluation of Cloud services. Based on this, we have selected a City Council in Portugal, whose CIO had doubts about what Cloud service he should purchase, which fits research problem. Due to the advantages of Cloud Computing, the CIO of the City Council (the DM in this case) wished to migrate some services, like e-mail and productivity, to the Cloud. However, he did not know how to choose the most adequate service alternative for the City

Council needs. Thereunto, we identified only two services that covered the City Council needs: Google Apps and Microsoft Office 365.

Table 2. Evaluation criteria with performance reference levels

Criteria	Performance Reference Levels	
	Good	**Neutral**
Client Support	The service provider has defined methods to support the client but is not able to communicate and report service failures	The service provider has no defined methods to support the client but is able to communicate and report service failures
Compliance with Standards	The service provider follows all the standards, processes and policies	The service provider follows some of the standards, processes, and policies
Data Ownership	90% of levels of rights	50% of levels of rights
Service Level Agreements Capacity	The service provider is able to negotiate all terms of the SLAs	The service provider is able to negotiate some terms of the SLAs
Adaptability to Client Requirements	The service provider is able to include core or important client requirements in the service	The service provider is able to include client requirements if they not require any modification in the service
Elasticity	100% of level of added resources	50% of level of added resources
Portability	The service can be ported to other service provider without disruption	The service can be ported to other service provider but can not move all the data
Availability	99% amount of time without interruptions per day	97% amount of time without interruptions per day
Maintainability	The service maintenance does not affect the service up time	The service maintenance stops the service
Reliability	The service can operate without failures under common unfavourable conditions (e.g. power failure)	The service can operate under unfavourable conditions but some components may not work
Risks	The service provider has an effective risk identification and treatment but no contingency plan	The service provider has o risk identification, no risk treatment, and no contingency plan
Acquisition and Transaction Cost	€0	€1000
Cost	€10	€20
Laws and Regulations	The service is subject to laws and regulations to protect clients against all kind of irregularities in the provider's country	The service is subject to laws and regulations only to protect clients against data losses in the provider's country
Innovation	The service is able to make all updates to new technologies and to include innovative features automatically	The service is able to make updates to new technologies but not automatically
Interoperability	The service is able to interact with other services	The service is able to interact only with services from the same service provider
Service Response Time	0,5 seconds	2 seconds
Confidentiality and Data Loss	The information is restricted to authorized people and a failure is promptly detected but no reported	The information is restricted to authorized people but there is no detection and reported failures
Data Integrity	The data stored is accurate and valid and backups are updated to the second	The data stored is accurate and valid and backups are updated monthly

A) Structuring the model. This first step began with some meetings with the City Council's DM in order to understand the decision context and to identify the evaluation criteria that should be used in the model. The DM validated all criteria listed in Table 2 and their descriptors of performance, accepting most of them as the essential criteria to their problem. However, the inclusion of "Acquisition and Transaction Cost" in model raised some doubts because in this kind of Cloud services there is a month for service experimenting without acquisition cost.

B) Evaluating the alternatives. During the second step we acted as a decision analyst guiding the decision process in order to help the DM. We used the M-MACBETH decision support system to display on the spot the model being developed.

The DM was asked to validate a neutral reference level on each criterion, which means to define a performance that would be neither positive nor negative in the linked objective, and a good reference level, which means a performance level considered significantly attractive in the light of the criterion.

Afterwards, more performance levels were added such that each criterion had at least three performance levels equally spaced. The DM was then asked to judge the differences in attractiveness between each two levels of performance, choosing one of the MACBETH semantic categories. **Figure 1** presents the DM's judgments matrix for the criterion "Availability". For example, the difference in attractiveness between 100% amount of time without interruptions per day and 99% was judged "weak", whereas the differences between 99% and 98%, 98% and 97%, 97% and 96% were deemed "moderate", which means that the DM values less the difference between 100% and 99% than the other mentioned differences.

	100	99	98	97	96
100	no	weak	moderate	strong	v. strong
99		no	moderate	strong	v. strong
98			no	moderate	strong
97				no	moderate
96					no

Figure 1. MACBETH judgements matrix for criterion "Availability"

The numerical scale was anchored on the two reference levels "neutral" and "good" to which were assigned the value scores 0 and 100, respectively. The M-MACBETH decision support system proposed a numerical scale based on the set of qualitative judgments inputted in the matrix of judgments using linear programming [8]. The proposed MACBETH scale was then subjected to DM analysis in terms of proportions of the resulting scale intervals. **Figure 2** presents the value function obtained for criterion "Availability".

Value functions were built in similar manner for the other 18 criteria. In some cases the DM decided to make "small" adjustments on the value scales.

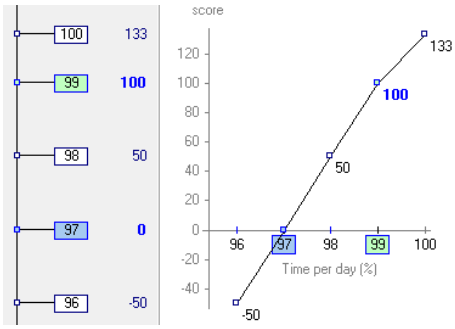

Figure 2. Value function for criterion "Availability"

Afterwards, the relative weights for the 19 criteria were assessed using the MACBETH weighting procedure. The DM was first asked to rank the criteria neutral-good swings by their overall attractiveness. We started by asking the question: "From the nineteen criteria, if you could choose just one criterion to move from a neutral performance to a good performance which objective would you choose?" The questioning procedure continued until the final ranking of neutral-good swings was achieved. Next, the DM was asked to judge the difference in attractiveness between each two neutral-good swings. With the DM's judgments inputted in the weighting matrix M-MACBETH generated the weights shown in **Figure 3**.

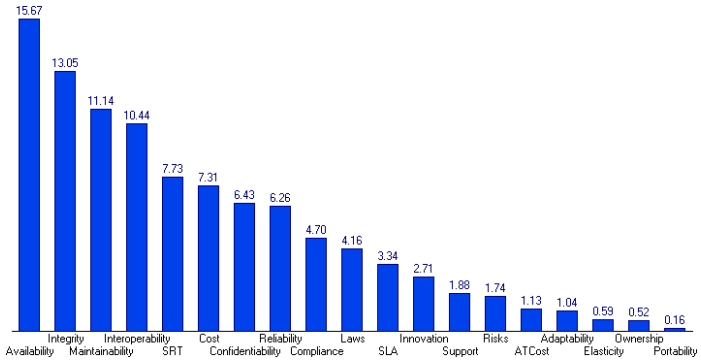

Figure 3. Weighting scale obtained for the evaluation criteria presented in Table 2

Then the DM validated the proposed MACBETH scale. For example, we was asked if the neutral-good swing on criterion "Integrity" is worth two times the neutral-good swing on criterion "Confidentiality and Data Loss", which the DM agreed.

C) Analysing the results. The performances of the Google Apps and Microsoft Office 365 upon each of the criteria were inputted in M-MACBETH. The software transformed the performances into the value scores, presented in **Figure 4**, using the value functions previously built, and calculated the overall scores for the alternatives (see column "Overall" in **Figure 4**). Google Apps ranked first with 102.08 overall units and Microsoft Office 365 ranked second with 81.21 overall units. Only Google Apps obtained an overall score higher than the score of the hypothetical alternative "Good at all" (i.e. a fictitious alternative that has a good performance in all the criteria), which shows that Google Apps is a very attractive alternative for the City Council. Microsoft Office 365 also is an attractive alternative, because its overall score is closer to the score of the hypothetical alternative "Good at all" than to the score of the hypothetical alternative "Neutral at all".

Options	Overall	Compliance	Ownership	SLA	Support	Adaptability	Elasticity	Portability	Availability	Maintainability
Google	102.08	100.00	125.00	-133.00	200.00	-100.00	37.00	0.00	129.70	100.00
[Good at all]	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Microsoft	81.21	100.00	125.00	-133.00	200.00	-100.00	37.00	0.00	129.70	0.00
[Neutral at all]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Weights :		0.0470	0.0052	0.0334	0.0188	0.0104	0.0059	0.0016	0.1567	0.1114

Reliability	Risks	ATCost	Cost	Laws	Innovation	Interoperability	SRT	Confidentiability	Integrity
233.00	100.00	100.00	160.00	150.00	100.00	100.00	67.00	0.00	100.00
100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100.00	200.00	100.00	77.00	150.00	0.00	0.00	67.00	250.00	100.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.0626	0.0174	0.0113	0.0731	0.0416	0.0271	0.1044	0.0773	0.0643	0.1305

Figure 4. Overall value scores of the alternatives

Observe in **Figure 4** that Microsoft Office 365 is better than Google Apps only in two criteria: "Risks" and "Confidentiality and Data Loss". A sensitivity analysis on the weight of criterion "Risks" showed that the weight of this criterion would need to be raised up from 1.74% to 18.7% to see Microsoft Office 365 be ranked first. A similar analysis showed that the weight of criterion "Confidentiality and Data Loss" would need to be raised up from 6.43% to 13.6% to see Microsoft Office 365 as the winner alternative. However, the DM did not consider plausible these changes on the weights.

Finally, a robustness analysis made with M-MACBETH considering simultaneous variations of ±3% on the weights of all criteria, though not allowing negative weights, revealed that Google Apps continues to be the most attractive of the two alternatives. A green cross in a cell in **Figure 5** means that the alternative in row, Google Apps, additively dominates the alternative in column, Microsoft Office 365, i.e., within these variations on the criteria weights Google Apps is always more attractive than Microsoft Office 365.

	Google	[Good at all]	Microsoft	[Neutral at all]
Google	=	?	+	+
[Good at all]	?	=	?	▲
Microsoft		?	=	+
[Neutral at all]				=

Figure 5. Robustness analysis

To conclude the process, we recommended to the City Council the selection of Google Apps, because it is the better alternative taking into account all the defined criteria and the judgments of preference made by the DM. In addition, the sensitivity and robustness analyses showed that Google Apps is a robust choice.

6 Evaluation

This section corresponds to the evaluation step of DSRM, which aim to observe and measure how well the artefact supports a solution of the problem. To make the evaluation we used the Moody and Shanks Quality Framework [11].

The Moody and Shanks Quality Framework is the result of research on how to evaluate and improve the quality of data models from the perspective of the multiple stakeholders and proposes eight quality factors [11]. We applied this framework to the demonstration by asking the DM about these eight quality factors and the results were the following:

- **Completeness**: the proposal is complete since the main criteria to evaluate Cloud services are present, and each DM can include or remove criteria and change their descriptors of performance to better assess criteria satisfaction.
- **Integrity**: there is no business rule or other constraint that prevents errors defining the criteria and their descriptors of performance of the proposal since it relies on interviews and observations.
- **Flexibility**: a DM can add or remove criteria to adjust the evaluation model to his organization's businesses and strategies.
- **Understandability**: the proposal is easy to understand since their language is close to the traditional usage in Cloud services but the DM do not know the decision analysis process and this phase is more difficult without a guide.
- **Correctness**: the proposal is correct and valid for their intentions.
- **Simplicity**: the proposal is simple to follow, and we verified that is simple to apply.
- **Integration**: the proposal is consistent with the problem and help organizations to make the best decision.
- **Implementability**: the proposal implementability is dependent on the law and policies in each organization. The City Council's CIO admitted to use it as an auxiliary tool.

The demonstration in the City Council allowed us to test our proposal in the research problem stated. The City Council suffered from the same problem, as we found in literature, and our proposal helped them to overcome it. The field case revealed that the method developed is a suitable tool for evaluating Cloud services, and the DM showed very interest in using it.

7 Conclusion

The research literature and publications from consulting enterprises consider that Cloud Computing has benefits, risks, challenges and issues. But all agree that organizations suffer when choosing which Cloud services they will contract, which reveals a generic and important problem: DMs are not prepared to evaluate Cloud services. As a result, they postpone the decision, then postponing the benefits of adopting Cloud services. We observed this problem in the concrete example of the City Council in which we demonstrated the proposed method.

In order to address this problem, we propose a method to evaluate Cloud services with an MCDA approach called MACBETH that simplifies the decision-making process in organizations adopting Cloud services. The proposal is based on the criteria defined in [31] and forces DM to value judgments in order to find out the most overall attractive Cloud service.

This paper has a particular focus on the multi-criteria evaluation process and its application to a City Council in Portugal, where two Cloud services (Google Apps and Microsoft Office 365) were evaluated. With this demonstration we conclude that our proposal is suitable and can be applied to evaluate Cloud services. The Moody and Shanks evaluation we performed supports this conclusion, as almost all quality factors were accomplished.

Regarding future work, more research effort related to the different Cloud models could be used in order to create criteria catalogues that could be applied to different Cloud models, such as SaaS, PaaS, and IaaS. In addition, our proposal can be further improved by developing a software tool specific for Cloud services evaluation.

References

1. Willcocks, L., Lacity, M.: The New IT Outsourcing Landscape – From Innovation to Cloud Services, Palgrave Macmillan (2012).
2. McAfee, A.: What Every CEO Needs to Know About the Cloud, MIT Sloan Management (2012).
3. IDC Portugal: Situação Actual e Tendências de Adopção de Serviços Cloud Computing em Portugal (in Portuguese), IDC Survey (2012).

4. Conway, G., Curry, E.: Managing Cloud Computing: A Life Cycle Approach, 2nd International Conference on Cloud Computing and Services Science (2012).
5. Belton, V., Stewart, T.J.: Multiple Criteria Decision Analysis: An Integrated Approach, Kluwer Academic Publishers (2002).
6. Bana e Costa, C.A., Vansnick, J.C.: The MACBETH approach: Basic Ideas, Software, and an Application, Advances in Decision Analysis, 4, 131-157, Kluwer Academic Publishers (1999).
7. Bana e Costa, C.A., De Corte, J.M., Vansnick, J.C.: MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique), Wiley Encyclopedia of Operations Research and Management Science, John Wiley & Sons, 4, 2945-2950 (2011).
8. Bana e Costa, C.A., De Corte, J.M., Vansnick, J.C.: MACBETH, International Journal of Information Technology & Decision Making, 11(2), 359-387 (2012).
9. Keeney, R. L., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs, John Wiley & Sons (1976).
10. von Winterfeldt, D., Edwards, W.: Decision Analysis and Behavioral Research, Cambridge University Press (1986).
11. Moody, D.L. and Shanks, G.G.: Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework, Information Systems, 28, 619-650 (2003).
12. Hevner, A.R., March, S.T., Park, J.: Design Science in Information Systems Research, MIS Quarterly, 28(1), 75-105 (2004).
13. Peffers, K., Tuunamen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research, Journal of Management Information Systems, 24(3), 45-77 (2008).
14. Weeden, S., Valiente, T.: Cloud Computing Every Silver Lining Has a Cloud, Citi Research, Tech. Rep. (2012).
15. Staten, J.: Is Cloud Computing Ready for the Enterprise?, Forrester Report (2008).
16. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: A View of Cloud Computing, Communications of the ACM, 53(4) (2010).
17. Mell, P., Grance, T.: The NIST Definition of Cloud Computing, National Institute of Standards and Technology, Tech. Rep. (2009).
18. Garg, S., Versteeg, S., Buyya, R.: A Framework for Ranking of Cloud Computing Services, Future Generation Computer Systems (2012).
19. Lacity, M., Willcocks, L.: Advanced Outsourcing Practice: Rethinking ITO, BPO and Cloud Services, Palgrave Macmillan (2012).
20. Rose, A., Alias, A.R., Azizah, H.N., Rozi, N.: The Performance Measurement Model for ICT Service Quality, 4th International Conference on Wireless Communication, Networking and Mobile Computing (WiCom), 1-5 (2008).
21. ISO/IEC: ISO/IEC 9125 – Information Technology – Software Product Evaluation – Quality Characteristics and Guidelines for Their Use (1991).
22. Sevcik, P.: Defining the Application Performance Index, Business Communication Review, 8-10 (2005)
23. Apdex Alliance Inc.: Application Performance Index – Apdex Technical Specification, Apdex Alliance, v. 1.1 (2007).
24. Hefley, W.E., Loesche, E.A.: eSourcing Capability Model for Client Organizations, Software Engineering Institute, Carnegie Mellon University (2006).
25. Cloud Services Measurement Initiative Consortium (CSMIC): Service Measurement Index, Carnegie Mellon University, Silicon Valley (2011).
26. Cloud Services Measurement Initiative Consortium (CSMIC): Cloud Services Measurement Initiative Consortium Overview, Carnegie Mellon University, Silicon Valley (2012).
27. Saaty, T.L.: The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill (1980).
28. Bana e Costa, C.A., De Corte, J.M., Vansnick, J.C: M-MACBETH Version 1.1 User's Guide. Retrieved from http://www.m-macbeth.com/ (2005).
29. Bana e Costa, C.A., Vansnick, J. C.: A Critical Analysis of the Eigenvalue Method used to Derive Priorities in AHP, European Journal of Operational Research, 187(3), 1422-1428 (2008).
30. Dyer, J.S.: Remarks on the Analytic Hierarchy Process, Management Science, 36(3), 249-258 (1990).
31. Costa, P., Santos, J., Mira da Silva, M.: Evaluation Criteria for Cloud Services, IEEE 6th International Conference on Cloud Computing (2013).