# Evaluating Cloud Services Using a Multiple Criteria Decision Analysis Approach

Pedro Costa[1], João Carlos Lourenço[2], and Miguel Mira da Silva[1]

[1] Department of Computer Science and Engineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[2] Centre for Management Studies of IST (CEG-IST), Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
{pedro.c.costa,joao.lourenco,mms}@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 services provider because there are no established guidelines to help make that decision. In order to address this problem, we propose a multi-criteria model to evaluate Cloud services using the MACBETH method. The proposed method was demonstrated in a City Council in Portugal to evaluate and compare two Cloud services: Google Apps and Microsoft Office 365.

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

## 1 Introduction

IT industry is evolving and there is a new business model, which is revolutionizing and changing it: Cloud services. Organizations can now contract services from the Cloud rather than owning the assets to provide those services [1-2]. However, despite the growing adoption of Cloud services, most decision-makers continue to express some concerns [3], because these services are still in their beginning and quite far from maturity. In fact, decision-makers have doubts about what, when, and how they should migrate to the Cloud, because there are no clear guidelines in this area [4]. In addition to this, decision-makers may not have the knowledge about the real benefits, risks, and costs associated with Cloud solutions, which may lead them to postpone the decision to migrate to Cloud. Therefore, organizations need a systematic tool 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].

In this paper we study how to help a decision-maker (DM) to evaluate Cloud solutions. To address this problem we propose a Multiple Criteria Decision Analysis (MCDA) approach [5], based on the MACBETH method [6-7], to build a multi-criteria value model [8-9] to evaluate Cloud services. Complementary, our proposal

S. Basu et al. (Eds.): ICSOC 2013, LNCS 8274, pp. 456–464, 2013.

should also: *(i)* clarify DMs doubts and fears about Cloud Computing; *(ii)* be easy to apply and not requiring specialized expertise; *(iii)* be able to provide understandable results; and *(iv)* be less expensive than current solutions.

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 and office) to the Cloud. 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 of these options, which depicted their overall attractiveness for the City Council. We used the feedback of the DM during the process and the Moody and Shanks Framework [10] to evaluate our proposal, which showed that it 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 [11]. The steps of DSRM that were used to organize the paper are: problem identification and motivation; objectives of a solution definition; design and development; demonstration; evaluation; and communication [12].

## 2 Related Work

A decision problem typically involves balancing multiple, and often conflicting, criteria. 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 section 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 correspond 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 options *a* and *b*, where $z_i(a) \geq z_i(b)$ for all criteria *i*, we can say that option *a* outranks option *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 options 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 options. AHP uses pairwise comparisons of options to score the options on each criterion and uses pairwise comparison of criteria to weight the criteria, assuming ratio scales for all judgments. The overall score of an option is obtained by the weighted summation of its scores on the different criteria [5], [13].

**MACBETH.** MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) is an approach for multi-criteria value measurement [8-9]. It uses semantic judgments about the differences in attractiveness of several stimuli to help a

DM quantify the relative attractiveness of each option. 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 [14] it automatically verifies their consistency. It subsequently generates a numerical scale, by solving a linear programming problem, which is representative of the DM's judgments. Through a similar process it permits the generation of weighting scales for criteria [7].

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 the options. AHP generates global scores to represent the overall preference upon the options, 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 [15-16]. A MACBETH advantage over other methods for multi-criteria value measurement is that it only requires qualitative judgments to score options and to weight criteria. Furthermore, its decision support system (M-MACBETH) is able to compute the overall value scores of the options by applying the additive model, and to make extensive sensitivity and robustness analysis.

## 3 Proposal

To address the problem specified in Section 1 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's 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. Two reference levels (e.g. "neutral" and "good") must be defined on each descriptor of performance. Then, other performance levels may be added to the descriptor, if needed. We created a template with the reference levels of performance for all Cloud services evaluation criteria presented in [17] (see Table 1). 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 Options.** In the second step the DM is asked about his preferences in order to build a value function for each criterion and to weight the criteria. To build a value function for a criterion 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 [7] to generate a numerical value scale compatible with the DM's judgments, which should be validated in terms of the proportions of the resulting scale intervals.

**Table 1.** Evaluation criteria with their respective reference levels

| Criteria | 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 meth ods to support the client but is able to communicate and report service failures |
| Compliance with Standards | The service provider follows all the stand ards, 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 Re-quirements | 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 unfavorable conditions (e.g. power failure) | The service can operate under unfavorable conditions but some components may not work |
| Risks | The service provider has an effective risk identification and treatment but no contin-gency plan | The service provider has o risk identifica-tion, no risk treatment, and no contingen-cy plan |
| Acquisition and Transaction Cost | €0 | €1000 |
| Cost | €10 | €20 |
| Laws and Regulations | The service is subject to laws and regula-tions to protect clients against all kind of irregularities in the provider's country | The service is subject to laws and regula-tions 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 innova-tive 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 |

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 options (factual data) are converted into value scores, using the value functions previously built for the criteria, and an overall value score is calculated for each option by weighted summation of its value scores. A final ranking of the options 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.

# 4 Demonstration

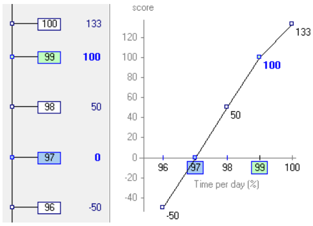
The main objective of this proposal is to construct a tool that enables any organization to evaluate Cloud services options. Based on this, we have selected a City Council in Portugal, whose CIO (the DM in this case) had doubts about what Cloud service he should purchase. Due to the advantages of Cloud Computing, the DM wished to migrate some services (e-mail and productivity) to the Cloud. However, he did not know how to choose the most adequate service option for the City Council. Only two services covered the City Council needs: Google Apps and Microsoft Office 365. We acted as a decision analyst guiding the decision process in order to help the DM. The M-MACBETH decision support system was used to display the model being developed.

**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 accepted all criteria listed in Table 1 as the essential criteria to their problem. Then the DM was asked to validate for each criterion a "neutral" reference level (i.e. a performance that would be neither positive nor negative in the linked criterion) and a "good" reference level (i.e. a performance level considered significantly attractive in the light of the criterion). For example, the "neutral" and "good" reference levels defined for the criterion "Availability" were 97% and 99%, respectively (Table 1). Afterwards, more performance levels were added such that each criterion had at least three performance levels equally spaced.

**B) Evaluating the Options.** A value function was built for each criterion by asking the DM to judge the differences in attractiveness between each two levels of performance, choosing one of the MACBETH semantic categories. Figure 1a presents the DM's judgments matrix for the criterion "Availability", where we can see, for example, that the difference in attractiveness between 100% and 99% amount of time without interruptions per day was judged "weak", whereas the differences between 99% and 98%, 98% and 97%, and 97% and 96% were deemed "moderate", which means that the DM values less the difference between 100% and 99% than the other mentioned differences. The numerical value scale was anchored on the 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 value scale based on the set of qualitative judgments inputted in the matrix of judgments using linear programming. The proposed MACBETH scale was then subjected to DM analysis in terms of proportions of the resulting scale intervals. Figure 1b presents the value function obtained for criterion "Availability". The value functions for the other 18 criteria were built in a similar manner.

| | 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 |

(a)




(b)

**Fig. 1.** MACBETH judgements matrix (a) and value function (b) 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 19 criteria, if you could choose just one to move from a neutral performance to a good performance which criterion would you select?" The DM's answer identified the criterion with the highest weight. 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 2. Then the DM validated the proposed MACBETH scale. For example, he 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.

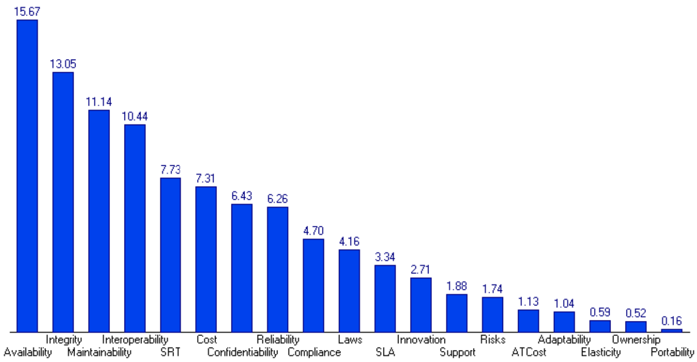



**Fig. 2.** Weighting scale obtained for the evaluation criteria

**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 3, using the value functions previously built, and calculated the overall scores for the options (see column “Overall” in Figure 3). 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 option “Good at all” (i.e. a fictitious options that has a good performance in all the criteria), which shows that Google Apps is a very attractive option for the City Council. Microsoft Office 365 also is an attractive option, because its overall score is closer to the score of the hypothetical option “Good at all” than to the score of the hypothetical option “Neutral at all”. Observe in Figure 3 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 increased 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 increased from 6.43% to 13.6% to see Microsoft Office 365 as the winner option. However, the DM did not consider plausible these changes on the weights. 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 continued to be the most attractive option.

| 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 |

| Options | Reliability | Risks | ATCost | Cost | Laws | Innovation | Interoperability | SRT | Confidentiability | Integrity |
|---|---|---|---|---|---|---|---|---|---|---|
| Google | 233.00 | 100.00 | 100.00 | 160.00 | 150.00 | 100.00 | 100.00 | 67.00 | 0.00 | 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 | 100.00 | 200.00 | 100.00 | 77.00 | 150.00 | 0.00 | 0.00 | 67.00 | 250.00 | 100.00 |
| [ Neutral at all ] | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Weights : | 0.0626 | 0.0174 | 0.0113 | 0.0731 | 0.0416 | 0.0271 | 0.1044 | 0.0773 | 0.0643 | 0.1305 |

**Fig. 3.** Overall value scores of the options

To conclude the process, we recommended to the City Council the selection of Google Apps, because it is the better option 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.

## 5 Evaluation

The Moody and Shanks Quality Framework propose eight quality factors to evaluate the quality of data models [10]. We applied this framework to the demonstration by asking the DM about these eight quality factors. The results were the following: *(i)* **completeness**: the proposal is complete since the main criteria to evaluate Cloud services are present; *(ii)* **integrity**: there is no business rule that prevents errors defining the criteria and their descriptors of performance of the proposal since it relies on interviews and observations; *(iii)* **flexibility**: a DM can add or remove criteria to adjust the evaluation model to his organization's businesses and strategies; *(iv)* **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; *(v)* **correctness**: the proposal is correct and valid for their intentions; *(vi)* **simplicity**: the proposal is simple to follow and we verified that is simple to apply; *(vii)* **integration**: the proposal is consistent with the problem and help organizations to make the best decision; and *(viii)* **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.

This demonstration 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.

## 6 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 would contract, which reveals a generic and important problem: typically, DMs are not prepared to evaluate Cloud services. To address this problem, we propose to evaluate Cloud services with an MCDA method called MACBETH that simplifies the decision-making process in organizations adopting Cloud services.

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, Basingstoke (2012)
2. McAfee, A.: What Every CEO Needs to Know About the Cloud. Harv. Bus. Rev. 89, 124–132 (2011)
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. In: 2nd International Conference on Cloud Computing and Services Science (CLOSER), Porto, Portugal (2012)
5. Belton, V., Stewart, T.J.: Multiple Criteria Decision Analysis: An Integrated Approach. Kluwer Academic Publishers, Boston (2002)
6. Bana e Costa, C.A., Vansnick, J.C.: The MACBETH approach: Basic Ideas, Software, and an Application. In: Meskens, N., Roubens, M.R. (eds.) Advances in Decision Analysis, vol. 4, pp. 131–157. Kluwer Academic Publishers, Dordrecht (1999)
7. Bana e Costa, C.A., De Corte, J.M., Vansnick, J.C.: MACBETH. Int. J. Inf. Technol. Decis. Mak. 11, 359–387 (2012)
8. Keeney, R.L., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley & Sons, New York (1976)
9. von Winterfeldt, D., Edwards, W.: Decision Analysis and Behavioral Research. Cambridge University Press, Cambridge (1986)
10. Moody, D.L., Shanks, G.G.: Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework. Inf. Syst. 28, 619–650 (2003)
11. Hevner, A.R., March, S.T., Park, J.: Design Science in Information Systems Research. MIS Q. 28, 75–105 (2004)
12. Peffers, K., Tuunamen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. J. Manage. Inf. Syst. 24, 45–77 (2008)
13. Saaty, T.L.: The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. McGraw-Hill (1980)
14. Bana e Costa, C.A., De Corte, J.M., Vansnick, J.C: M-MACBETH Version 1.1 User's Guide (2005), `http://www.m-macbeth.com/` (retrieved)
15. Bana e Costa, C.A., Vansnick, J.C.: A Critical Analysis of the Eigenvalue Method used to Derive Priorities in AHP. Eur. J. Oper. Res. 187, 1422–1428 (2008)
16. Dyer, J.S.: Remarks on the Analytic Hierarchy Process. Manage. Sci. 36, 249–258 (1990)
17. Costa, P., Santos, J., Mira da Silva, M.: Evaluation Criteria for Cloud Services. In: IEEE 6th International Conference on Cloud Computing, Santa Clara Marriott, CA, USA (2013)