Price Modeling of IaaS Providers
An Approach Focused on Enterprise Application Integration

Cássio L. M. Belusso[1], Sandro Sawicki[2], Vitor Basto-Fernandes[3], Rafael Z. Frantz[2] and Fabricia Roos-Frantz[2]

[1]*Federal University of Fronteira Sul, Cerro Largo, Brazil*

[2]*Department of Exact Sciences and Engineering, UNIJUÍ University, Ijuí, Brazil*

[3]*Department of Information Science and Technology, University Institute of Lisbon, Lisbon, Portugal*

Keywords: Cloud Computing, Enterprise Application Integration, Price Modeling, Linear Regression, IaaS.

Abstract: One of the main advances in information technology today is cloud computing. It is a great alternative for users to reduce costs related to the need to acquire and maintain computational infrastructure to develop, implement and execute software applications. Cloud computing services are offered by providers and can be classified into three main modalities: Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS) and Infrastructure-as-a-Service (IaaS). In IaaS, the user has a virtual machine at their disposal with the desired computational resources at a given cost. Generally, the providers offer infrastructure services divided into instances, with pre-established configurations. The main challenge faced by companies is to choose the instance that best fits their needs among the many options offered by providers. Frequently, these companies need a large computational infrastructure to manage and improve their business processes and, due to the high cost of maintaining local infrastructure, they have begun to migrate applications to the cloud in order to reduce these costs. In this paper, we introduce a proposal for price modeling of instances of virtual machines using linear regression. This approach analyzes a set of simplified hypotheses considering the following providers: Amazon EC2, Google Compute Engine and Microsoft Windows Azure.

1 INTRODUCTION

Cloud computing is a significant innovation in the field of information technology as well as one of the greatest perspectives for growth in the coming years. Despite having gained much attention recently, the concepts involved in this new computing resource have been in development for a long time. However, there is not yet a formal definition for the concept of cloud computing. Essentially, it is the utilization of the most diverse applications over the Internet, in the same way as if they were installed on a computer or other physical device.

An important advantage of cloud computing is its cost-benefit relation, where the user pays only for what is used (Murthy et al., 2012). In addition, users of services in the cloud are provided with maintenance services, upgrades, backup, security and other resources that would be needed were the applications installed on their own computer.

Cloud computing services are offered by providers and there are several on the market offering various services, with emphasis on software services (SaaS - Software-as-a-Service), where certain softwares are provided and paid for on a per-use basis; platform services (PaaS - Platform-as-a-Service), where the user is provided with an environment to design, test, and deploy custom applications; and infrastructure services (IaaS - Infrastructure-as-a-Service), where the user is provided with virtual machines and manages resources as desired (CPU, memory, storage, data transfers, network bandwidth, etc). Our research focuses specifically on IaaS.

IaaS providers offer a range of different plans for certain instances, which vary according to the needs of each client. Generally, individuals and even small businesses require instances comprising small demands. Large companies, however, have large demands, and thus have been mainly responsible for more substantial investments in cloud computing.

One of the problems faced by users is that the cost of cloud computing resources varies not only between the numerous providers, but also between the different instances that each offers. This hinders decision

Belusso, C., Sawicki, S., Basto-Fernandes, V., Frantz, R. and Roos-Frantz, F.
Price Modeling of IaaS Providers - An Approach Focused on Enterprise Application Integration.
DOI: 10.5220/0006371603710376
In *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017) - Volume 1*, pages 371-376
ISBN: 978-989-758-247-9

making, because often the lowest price is not always the best choice, since the user may be looking for a provider/instance that offers better Quality-of-Service (QoS). For example, if users acquire an instance with more resources than are needed, they will be spending money unnecessarily.

Instance prices charged by providers are based on the amount of computational resources employed by users. However, an important question attracting the attention of research community is related to how these prices are set. Whereas some research shows that there is a mechanism defined for this (Murthy et al., 2012; Kihal et al., 2012; Mitropoulou et al., 2016), others show that price is much more complex than was previously thought and that it is influenced by economic policy and the well-known law of supply and demand (Al-Roomi et al., 2013; Mazrekaj et al., 2016). After prices are set by providers, the final value of the instances can still change based on other factors, such as the location where the virtual machine is hosted or the operating system chosen by the user.

In the context of Enterprise Application Integration (EAI), cloud computing provides a high-capacity computing infrastructure at a low cost, in which integration solutions can be deployed and run. Integration solutions are extremely important in the process of integration, because they are softwares that act as a communication link between the different applications contained in the software ecosystem of companies, allowing sharing information across them quickly and efficiently (Frantz et al., 2016). Figure 1 illustrates the basic scheme of an integration solution.

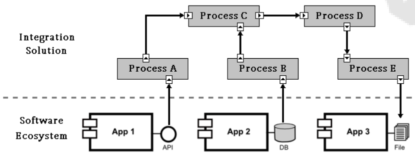

Figure 1: Integration solution.

Despite large companies having made significant investments in cloud computing to improve their business processes, the providers do not offer a method to describe the variability of the services and the constraints among them, so that these models can be used in the decision-making process (Hernández et al., 2015). This study presents a new proposal for modeling the instance prices charged using linear regression by three of the main providers of cloud computing: Amazon EC2 (Amazon, 2016), Google Compute Engine (Google, 2016) and Microsoft Windows Azure (Azure, 2016). Our aim is that this new model can contribute to decision-making related to which provider/instance best adapts to the needs of the companies, resulting in a significant reduction of time and money from the deployment and implementation of integration solutions in the cloud. The proposal for price modeling of instances of virtual machines is described by presenting some of the simplifying assumptions adopted in this new model.

The remainder of this article is organized as follows: Section 2 summarizes some papers that present price modeling proposals for some providers on the market, as well as some prospects for analysis of price policy practiced by them; Section 3 presents a study of the hypotheses for the development of the new proposal and theories chosen for the construction of the model; Finally, in Section 4 the conclusions and some prospects for future studies are presented.

2 RELATED WORK

The studies presented in this section seek to understand how IaaS providers establish their pricing policy for their services. It is widely understood that this is not an easy task, which explains why there are different approaches adopted.

Murthy et al. (2012) present a comparative study of pricing and billing models of some IaaS providers. The authors consider storage space as the main hardware requirement, but CPU and memory are also used to compare the instances. They also analyze differences in the prices of instances for two different operating systems and in the different forms of purchasing or contracting the service. On the other hand, the hosting of virtual machines at different geographical locations is not considered, even though this can have a great influence on prices.

Rodamilans (2014) states that the problem of selection of instances can be solved by characterizing them and the applications, considering the financial cost and the performance of the virtual machine. He proposes a method for selecting cloud providers using the following steps: characterization, selection and implementation. For this purpose, it is not considered necessary to run the application in all instances to decide which is the most appropriate, rather it is only necessary to identify the dominant computational resource.

Mazrekaj et al. (2016) highlight that the objective of the providers in offering cloud computing services is always to get the highest revenue through their pricing schemes, while users seek the highest QoS at a low price. Thus, they make a price comparison of some models and pricing schemes of cloud services providers based on the following aspects: service, quality, fair price and importance in the market.

Chun and Choi (2014) analyze two pricing schemes practiced by providers of cloud computing: *subscription* and *pay-per-use*. However, unlike most proposals studied, the authors make an analysis from the perspective of the provider. They claim that the different pricing schemes are a consequence of the providers' different cost structures, which entail everything from the investment in cloud infrastructure to the cost of a service being idle for a certain period of time.

Kihal et al. (2012) present two methods to improve the transparency of the prices charged by providers of IaaS. The first is the hedonic pricing method, in which the fraction that each component contributes to the final cost of the instance is calculated. In this method, the instances need to be analyzed separately, making it impossible to compare providers. To work around this problem, the authors developed the Pricing Plan Comparison (PriCo), a method that identifies the best profile for each provider. In both cases, the authors considered as the main hypothesis that customers seek the cheapest service, not considering, for example, the QoS. During the analysis, the IaaS components memory, CPU and storage, Windows operating system and *On-demand* instances were used.

Mitropoulou et al. (2016) propose a price index based on the hedonic pricing method, which accounts for different factors in pricing models of cloud computing services at the IaaS level, including the geographic location of the provider. It is a mathematical approach, specifically of regression models (linear, exponential, for comparison), where the price of a service is related to its characteristics. Using the hedonic method, the authors allowed for an adjustment of the price of a service for the quality, not the quantity. The authors collected data from providers and analyzed in which regions they have hosted virtual machines. The model also accounts for qualitative variables related to the operating system and the purchasing or contractual model. Quantitative variables include amount of memory, CPU, storage and transfer_out. The authors did not consider some factors in the form of model variables because, according to them, they did not have much influence on the final price of the instances.

Our review of the latest research shows that there are no proposals directly related to selection of IaaS providers for migration of integration solutions to the cloud. Likewise, instruments or methodologies capable of quantifying the demand for computational resources that an integration solution consumes were not found. In this sense, our model can provide a needed new approach.

3 PROPOSED MODELING

In this paper, we present a proposal for modeling prices adopted by IaaS providers. The main focus is on Enterprise Application Integration, specifically in integration solution migration to the cloud. Due to this peculiarity, some assumptions need to be established and studied in this first stage, considering the aims and target audience.

The objective assumed in modeling is to always obtain the instance with the lowest price, but with a QoS capable of performing the predetermined computational demand. In addition, some factors influencing prices are analyzed, as well as the need to build them into the model.

3.1 Simplifying Hypotheses

Among the many IaaS providers on the market, in this study we opted to analyze three of them: Amazon EC2, Google Compute Engine and Microsoft Windows Azure. This choice was based on their popularity and, in the case of Amazon EC2, also due to the complexity of instance pricing. Some proposals are able to analyze more than one provider at the same time, and this is considered to be a *Cross-Provider* comparison (Hernández et al., 2015).

For this purpose, simplifying hypotheses are needed, initially, for some factors that influence, directly or indirectly, the prices charged by providers. These include:

- *The Operating System:* opting for a virtual machine that uses the Linux operating system as opposed to Windows, for example, can result in a significant difference in the price of the instances. For some providers, the price may also vary if the platforms are different. Information collected directly from the Internet homepages of providers indicates that an instance is always cheaper when opting for Linux.
- *Instance Type:* depending on the requirements for the deployment and implementation of integration solutions, instances with smaller or larger configurations, or instances of a *High-Memory* or *High-CPU* type, may be chosen, all of them grouped according to particular hardware requirements. *High-Memory* instances have large amounts of memory, while the amount of other components of the instance is small or has small variations. Similarly, *High-CPU* instances have a large processing capacity. Application integration solutions are light and don't require a lot of storage space. Thus, this requirement is not considered. Some providers have instances that differ

with regard to the storage unit, which can be Hard Disk Drive (HDD) or Solid-State Drive (SSD), and this can lead to changes in price. There is also grouping of instances considering the users' profiles. For those who use little, providers offer instances with a smaller configurations, while the instances with more powerful configurations are used by large companies, which have a large volume of applications running simultaneously.

- *Discount Model:* generally, the providers offer this to users who acquire an instance for a given amount of time. The longer the time of acquired, the greater the discount will be for the usage rate. There are also cases where the user can pay a percentage of the total cost in advance, where the higher the percentage of this payment, the greater the discount obtained. Other forms of discount include that for allocation of virtual machines in specific regions (Amazon, 2016), for percentage of monthly use of the virtual machine (Google, 2016) and for instances with cost above a pre-set value (Azure, 2016). Due to the different types of discounts offered, the elaboration of a model that includes all of them is a complex task, considering the peculiarities of each. For example, consider a provider *P*, for which one of the discount modes is given for the amount paid, and two clients, *A* and *B*, which have spent the same amount of money and, so, theoretically would get the same discount. Assuming that client *A* has been using the service provider *P* for a long time while client *B* is using the service for the first time. Customer *A* may be offered a discount greater due to their allegiance, thus partaking of another form of discount. Therefore, considering this imminent difficulty in modeling discounts, they are not considered in the proposed model.
- *Geographical Location of the Provider:* many providers have virtual machines that are hosted in various parts of the world and the price of the instance can change considerably within the same provider. Some providers have more than 10 possible locations (Amazon, 2016; Azure, 2016), making it difficult to compare. However, users may not always be able to choose the location in which the instance is cheaper, because there may be legal issues of the country in which the virtual machine is hosted. Mitropoulou et al. (2016) analyzed the influence of geographical location on the prices charged by providers and choose to group the different regions covered by continent. In the present proposal, which focuses on the lowest price with an ideal QoS, an adjacent survey of each of the providers should be made in order to identify regions with the lowest prices, limited to, at most, five of them.
- *Billing Model:* in addition to having numerous instances composed of different configurations, the providers also offer different pricing models. For example, Amazon EC2 provides three pricing models: *On-demand*, *Reserved* and *Spot*. For *On-demand* instances, the user pays according to use, without a long-term commitment; in *Reserved* instances, the user acquires the instance for a given amount of time and, because of this, pays a lower usage rate (per hour or per minute); finally, *Spot* instances allow the user to take part of a kind of auction for unused computing capacity, which can generate savings of up to 90 % in relation to the *On-demand* mode. However, the spot price fluctuates based on supply and demand for available capacity. If the user's offer is over that of the spot price, the instance will continue running; otherwise, the service will be interrupted. The need to stand out on the cloud computing market makes providers offer different pricing models, among which the *On-demand* model and the *Reserved* model are the most common. The *Reserved* models are always more economical in the long term and the price tends to fall even more as the subscription time increases (an analysis of pricing models Amazon EC2 can be found at Murthy et al. (2012)).
- *Period of Use:* finally, a more detailed study seeks to identify whether all of the chosen providers change their prices for the use of the services at different times and whether this alteration is significant. The user may choose to run applications on different schedules to achieve a reduction in the cost of their virtual machine. At peak hours, the price of the instances tends to be larger or the discount offered, if any, may be lower. Each provider defines peak hours.

3.2 Mathematical Model

Multiple linear regression was used to estimate the price of the instances for the three selected providers, due to the linear behavior of the data collected from the providers. During data collection, all the possibilities of geographical location of each of the providers were considered, as well as information pertaining to Linux and Windows operating systems. The pricing model considered was the *On-demand* model.

Using the regression method, it is possible to estimate the cost of individual characteristics that influence the final price of the instances by calculating the coefficients for each of the variables of the model.

The multiple linear regression model adapted to this proposal is described by Equation 1.

$$C_i = a_0 + a_1 X_{i1} + a_2 X_{i2} + ... + a_n X_{in} + \varepsilon_i, i = 1, ..., n \quad (1)$$

where,

- i refers to the i-th instance of any one of the three chosen providers;
- n is the number of variables of the model;
- C_i is the final price of the instance i;
- a_i are the coefficients of regression to be calculated, i.e. the share of participation of feature X in the final price of the instance i;
- X_{in} are the independent variables of the model, that is each of the characteristics that influence the final price of the instance i;
- ε_i is the residual error of the regression for each instance i.

The model variables can be summarized as quantitative variables, composed by hardware requirements (for example, CPU, memory and storage), and qualitative variables, consisting of software requirements (for example, operating system and platform) and other variables (for example, geographical location). Qualitative variables of the model are represented using Dummy variables, where a certain trait can take on the value 0 or 1 (Wonnacott and Wonnacott, 1990).

The application of the model can be better understood from the following example:

Example: an integration solution requires 4 GB of RAM, 2 processing cores and 200 MB of storage to be executed. By inserting this demand into the model, the price of all instances able to perform this solution is calculated and the cheapest instance is indicated. The Dummy variables are defined and inserted previously into the model. For the operating system, for example, the value of 1 for Windows and 0 for Linux can be set. If the option is Windows, multiplying the coefficient from the regression model for 1; being Linux, the coefficient is multiplied by 0. Assuming that an instance is more expensive when you choose Windows, the coefficient of this variable should be positive, generating an increase in the final price of the instance; in the case of the instance being cheaper, the coefficient should be negative, reducing the price. The same process can be adopted for the other qualitative variables. The coefficients of the quantitative variables represent the unit price of each of the hardware components, i.e. the cost per unit of RAM, CPU and storage.

Because it is a method of estimation, the level of reliability of the linear model must be considered. In addition, most models feature non-linearity and, because of that, a non-linear model may be more adequate. However, in these cases, due to difficulties encountered in non-linear modeling, the alternative used by researchers is the linearization of the variables, which sometimes does not cause major distortions to the results.

4 CONCLUSIONS

This study proposed a scheme for modeling the instance prices charged by providers of cloud computing at the IaaS level, using a linear regression model and some simplifying hypotheses. This proposal considered the providers Amazon EC2, Google Compute Engine and Microsoft Windows Azure.

The literature review show that no proposal to date has been geared toward implementation and execution of integration solutions in the cloud. The absence of a mechanism that can offer companies the opportunity to compare the services offered by providers makes this a new and promising approach, considering that, currently, all information must be collected directly from the Internet homepage of providers, delaying the decision-making process and, consequently, making this task quite onerous.

For future studies, it is necessary to first obtain the coefficients of the regression model and verify their level of reliability. Complementarily, it is possible to build a method of estimation able to quantify the demand for computational infrastructure necessary to perform an integration solution.

ACKNOWLEDGEMENTS

The research work on which we report in this paper is supported by CNPq, CAPES, FAPERGS, and the internal Research Programme at UNIJUI University. First author is also thanks the Federal University of Fronteira Sul (UFFS) for the support to the development of his research.

REFERENCES

Al-Roomi, M., Al-Ebrahim, S., Buqrais, S., and Ahmad, I. (2013). Cloud computing pricing models: a survey. *International Journal of Grid and Distributed Computing*, 6(5):93–106.

Amazon (2016). Amazon web services. https://aws.amazon.com/. Acesso em: 28/12/2016.

Azure (2016). Microsoft windows azure. https://azure.com/. Acesso em: 24/12/2016.

Chun, S. and Choi, B. (2014). Service models and pricing schemes for cloud computing. *Cluster Computing*, 17(2):529–535.

Frantz, R. Z., Corchuelo, R., and Roos-Frantz, F. (2016). On the design of a maintainable software development kit to implement integration solutions. *The Journal of Systems and Software*, 111:89–104.

Google (2016). Google cloud plataform. https://cloud.google.com/. Acesso em: 25/12/2016.

Hernández, I., Sawicki, S., Roos-Frantz, F., and Frantz, R. Z. (2015). Cloud configuration modelling: a literature review from an application integration deployment perspective. *Procedia Computer Science*, 64:977–983.

Kihal, S. E., Schlereth, C., and Skiera, B. (2012). Price comparasion for Infrastructure-as-a-Service. In *European Conference on Information Systems*. ResearchGate.

Mazrekaj, A., Shabani, I., and Sejdiu, B. (2016). Pricing schemes in cloud computing: an overview. *International Journal of Advanced Computer Science and Application*, 7(2):80–86.

Mitropoulou, P., Filiopoulou, E., Michalakelis, C., and Nikolaidou, M. (2016). Pricing cloud IaaS services based on a hedonic price index. *Computing*, pages 1–15.

Murthy, M. K. M., Sanjay, H. A., and Padmanabha, A. J. (2012). Pricing models and pricing schemes of IaaS providers: a comparison study. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, pages 143–147. ACM.

Rodamilans, C. B. (2014). *Uma metodologia para caracterização de aplicações e de instâncias de máquinas virtuais no ambiente de computação em nuvem*. Tese de doutorado, Escola Politécnica da Universidade de São Paulo, São Paulo.

Wonnacott, T. H. and Wonnacott, R. J. (1990). *Introductory Statistics for Business and Economics*. John Wiley & Sons, New York, 4th edition.