Journal of Network and Computer Applications 57 (2015) 102–118

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

A Petri net-based decision-making framework for assessing cloud services adoption: The use of spot instances for cost reduction

Maristella Ribas [1,a,*], C.G. Furtado [2,b], José Neuman de Souza [3,c], Giovanni Cordeiro Barroso [4,c], Antão Moura [5,d], Alberto S. Lima [6,c], Flávio R.C. Sousa [7,c]

[a] Techne Engenharia e Sistemas, São Paulo, Brazil
[b] Federal Institute of Ceará (IFCE), Fortaleza, Brazil
[c] Federal University of Ceará, Fortaleza, Brazil
[d] Federal University of Campina Grande, C. Grande, Brazil

ARTICLE INFO

Article history:
Received 20 February 2015
Received in revised form
25 May 2015
Accepted 3 July 2015
Available online 13 July 2015

Keywords:
Cloud computing
Spot instances
BDIM
AHP
Petri nets

ABSTRACT

Cloud services are widely used nowadays, especially in Infrastructure as a service (IaaS), with vendors offering several purchasing options and expanding the range of services offered on almost a daily basis. Cost reduction is a major factor promoting the adoption of cloud services among enterprises. However, qualitative factors need to be evaluated as well, thus rendering the decision regarding the adoption of cloud services among enterprises a non-trivial task for Information Technology (IT) managers. In this paper, we propose a *place/transition or* Petri net-based multi-criteria decision-making (MCDM) framework to assess a cloud service in comparison with a similar on-premises service. The framework helps IT managers choose between two such options, and can be used for any type of cloud service: Infrastructure as a Service (IaaS), Platform as a service (PaaS), Software as a service (SaaS), etc. Because its low cost is among the most important reasons for adopting cloud services, we also propose a Petri net to model cost savings using the spot instances purchasing option in public clouds. Through simulation of several scenarios, we conclude that spot instances present a very interesting cost-saving option in the auto-scaling process, even for simple business applications using few servers.

1. Introduction

Cloud computing is defined by the National Institute of Standard and Technology (National Institute of Standards and Technology, 2009) as a "model for enabling on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Infrastructure as a service (IaaS), Platform as a service (PaaS), and Software as a service (SaaS) are the classical categories of cloud services, although there are other proposed categories such as Database as a Service (DBaaS), Cache as a Service (CaaS), Unified Communications as a Service (UCaaS). In fact, cloud services are becoming so popular that some authors mention the category Everything as a Service (XaaS).

One of the many difficult decisions facing Information Technology (IT) managers nowadays is to choose between adopting an IT service in its cloud model and having the service hosted according to the conventional, on-premises model. The manager has to consider a large number of criteria, with cost being possibly the most important one. Cloud services are attractive because of the low-initial investment involved. However, if a cloud service is adopted by an organization, its IT department will no longer have any control over the infrastructure, and will completely rely upon the fulfillment of the Service-level Agreement (SLA) clauses in the contract to obtain the required service.

In this paper, we aim to shed some light on the business management problem of identifying the factors and parameters to determine the advantages and disadvantages of using cloud services in comparison with on-premises solutions for enterprise business applications. Furthermore, we focus on modeling and

* Corresponding author.
E-mail addresses: mari@techne.com.br (M. Ribas), cjunior@ifce.edu.br (C.G. Furtado), neuman@ufc.br (J.N. de Souza), gcb@fisica.ufc.br (G.C. Barroso), antao@dsc.ufcg.edu.br (A. Moura), albertosampaio@ufc.br (A.S. Lima), flaviosousa@ufc.br (F.R.C. Sousa).
[1] Rua Francisco Teixeira Alcantara, 500, Fortaleza CEP 60182360, Brazil.
[2] Rua Juvenal Galeno, 722, Benfica CEP 60015-340, Brazil.
[3] UFC-DC-Campus do Pici, Bloco 910, 60440-504 Fortaleza, CE, Brazil.
[4] Rua Princesa Isabel, 1618/101, Fortaleza CEP 60.015-061, CE, Brazil.
[5] Av. Aprigio Velloso 882 Bloco CN Sala 210 CEP 58429-140 Campina Grande, PB, Brazil.
[6] Rua Bento Albuquerque, 550 Apto 300, Coco, Fortaleza Cep 60192-060, –Ceara, Brazil.
[7] Rua Bento Albuquerque, 550 Apto 550, Coco, Fortaleza Cep 60192-060, –Ceara, Brazil.

http://dx.doi.org/10.1016/j.jnca.2015.07.002
1084-8045/

reducing the cost of the elasticity of cloud services. Also known as dynamic provisioning, elasticity "has become one of the most important features of a cloud computing platform" (Han et al., 2014). By using this feature, application owners can scale up or scale down resources used based on the computational demands of their applications, and need to pay only for the resources they actually use. Elasticity poses new challenges to resource management, as pointed in (Manvi and Shyam, 2014), and makes it more difficult to estimate cost, thus contributing to the greater complexity of the decision-making process.

As an example of the above-mentioned decision problem, consider a company that is developing a new platform to offer cloud services for developers of enterprise applications in a PaaS model. Developers will use PaaS to build and publish their applications. The PaaS provider will be required to host all applications, and will need the corresponding infrastructure. This infrastructure can be leased from public clouds, fully hosted on premises, or supplied in a hybrid manner, partly on-premises and partly rented from public clouds. The IT manager must determine the services that the company will host internally and those that will be hosted on public clouds in the most cost-efficient manner, all the while ensuring an acceptable quality of service for PaaS clients. There are instances of commercial PaaS that internally host their infrastructure (Google, AWS) as well as others that lease public IaaS (Heroku). Hence, the choice of service can be a very difficult decision for the IT manager.

In a previous study (Ribas et al., 2014), we proposed an initial version of a framework to support decisions on whether to lease cloud services by considering factors related only to SaaS. This initial framework combined several of the most relevant factors for decisions of this type (according to the literature) and provided an assessment of both options (public clouds and on-premises hosting). The framework was designed using Petri nets (PNs) in order to benefit from their formal description and provide a visual interface that is simple, yet powerful enough to perform simulations for different purposes. In this paper, we extend the framework to capture cost elements for any type of cloud service, and propose a Petri net model to simulate IaaS utilization and compute cost savings in several scenarios. The PN model used to estimate cost savings uses a particular purchasing option for virtual machines, called spot instances (Amazon, 2014). This purchasing option is currently supported by Amazon Web Services (AWS), the leader in the public IaaS market according to the Gartner Group (2014). Spot instances work in exactly the same manner as any other Amazon Elastic Compute Cloud (EC2) virtual machine. The difference lies only in the price scheme: the hourly price for spot instances is not fixed, and clients bid on how much they are willing to pay. AWS dynamically defines spot price, which varies in real-time based on supply and demand. If a client's bid is higher than the spot price, the spot instance commences. If the spot price changes and rises above the client's bid, the instance is terminated by AWS. In this paper, we refer to any type of virtual machine that can be rented in a public cloud as an "instance." Our contributions to the literature are constituted by: 1) Our proposed framework, where we selected, using recent literature as well as practical experience, important factors to be considered as well as a reliable method to assess all factors and provide a simple final score. Moreover, the diagram of the Petri net makes it very easy to understand the underlying methodology. 2) The results of simulation scenarios executed while investigating cost reduction through spot prices, where a discount of up to 60% can be obtained without a significant effort, simply by efficiently using cloud resource purchasing options.

The rest of this paper is organized as follows: Section 2 is devoted to a review of related work in the literature, which provides the parameters need to build the decision-making model. In Section 3, we present and discuss the PN model that we constructed to implement our decision-making framework, and provide an example in Section 4 to explicate our decision-making model. Section 5 contains an introduction to our PN models built to investigate cost reduction using spot prices, whereas Section 6 contains a description of our simulation scenarios, their results, and our discussion. We offer our conclusions and directions for future research in Section 7.

2. Related work

The literature on cloud computing is growing as cloud services are becoming more popular. For a systematic literature review, we searched for the terms "SaaS" and "evaluation," "evaluating," or "evaluate" among publications after 2011 on IEEE explore, the ACM digital library, and Science Direct (Elsevier). From the 228 items yielded by the search, we selected 32 papers that seemed relevant to our study. Many of these focused on the evaluation of SaaS strictly from the perspective of technical performance. However, our work follows research on business-driven IT management (BDIM) (Sauvé et al., 2006), which includes technical as well as business-centric views.

A few studies have assessed cloud solutions from the point of view of BDIM. The framework proposed in Sripanidkulchai and Sujichantararat (2012) is useful for comparing the cost of IaaS with that of on-premises datacenters. The primitive cognitive network process approach presented in Yuen (2012) is useful for selecting an offer of service (SaaS) from a list of providers of the same software. The survey in Benlian and Hess (2011) revealed that cost advantages were the strongest and the most consistent opportunity factor significantly affecting perceived opportunities in SaaS adoption, whereas security risks were the dominant risk factors, followed by performance and economic risks.

The study in Lee et al. (2013) presented results of a survey conducted in Korea to assess the adoption of SaaS and its related benefit to business, thus confirming the premises of the balanced scorecard (BSC) (Kaplan and Norton, 1996). The research in (Zardari and Bahsoon, 2011) proposed a process based on goal-oriented requirements engineering (GORE) to provide a systematic approach to evaluate a cloud provider. In Marston et al. (2011), the authors have provided SWOT analysis for the cloud computing industry, as well as various issues that will affect stakeholders.

Wu (2011a and 2011b) attempted to develop an explorative model that examined important factors affecting SaaS adoption, such as integrating Technology Acceptance Model (TAM)-related theories with additional imperative constructs, such as marketing effort, security, and trust. Security is the major risk affecting SaaS adoption according to most researchers (Benlian and Hess, 2011; Wu, 2011a, 2011b; Wu et al., 2011; Bayrak, 2013), whereas cost reduction is the major expected benefit (Benlian and Hess, 2011; Wu, 2011a, 2011b; Gupta et al., 2013; Bayrak, 2013).

Other interesting studies of cloud evaluation and adoption relate to pricing schemes (Rohitratana and Altmann, 2012), facets of security in the cloud (Boampong and Wahsheh, 2012), selection of cloud providers based on security, and privacy requirements (Mouratidisa et al., 2013). The cost analysis of on-premises solutions against SaaS solutions was conducted in detail in Bibi et al. (2012). In this study, unlike in our framework, the authors did not include qualitative benefits and risks to obtain a final score. Garg et al. (2013) used the analytic hierarchy process (AHP) to combine quality of service (QoS) attributes in order to address the problem of selecting a cloud provider, which differs from our problem of comparing on-premises solutions with cloud solutions.

We also conducted a search for "cloud cost model" and selected 43 papers relevant to our study. An interesting comparison of