Future Generation Computer Systems 41 (2014) 79–90

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Cost model based service placement in federated hybrid clouds

Jörn Altmann*, Mohammad Mahdi Kashef

Technology Management, Economics, and Policy Program, Department of Industrial Engineering, College of Engineering, Seoul National University, Seoul, South Korea

HIGHLIGHTS

- Comparison of cost models for cloud computing.
- Design of a comprehensive cost model for federated hybrid clouds.
- Design of an algorithm for service placements in federated hybrid clouds.
- Performed a sensitivity analysis with the help of the service placement algorithm.

ARTICLE INFO

Article history:
Received 11 September 2012
Received in revised form
17 August 2014
Accepted 19 August 2014
Available online 3 September 2014

Keywords:
Cloud federation
Service migration
Information technology cost factors
OPEX and CAPEX
Cloud economics
Cost model comparison

ABSTRACT

As cloud federation allows companies in need of computational resources to use computational resources hosted by different cloud providers, it reduces the cost of IT infrastructure by lowering capital and operational expenses. This is the result of economies of scale and the possibility for organizations to purchase just as much computing and storage resources as needed whenever needed. However, a clear specification of cost savings requires a detailed specification of the costs incurred. Although there are some efforts to define cost models for clouds, the need for a comprehensive cost model, which covers all cost factors and types of clouds, is undeniable. In this paper, we cover this gap by suggesting a cost model for the most general form of a cloud, namely federated hybrid clouds. This type of cloud is composed of a private cloud and a number of interoperable public clouds. The proposed cost model is applied within a cost minimization algorithm for making service placement decisions in clouds. We demonstrate the workings of our cost model and service placement algorithm within a specific cloud scenario. Our results show that the service placement algorithm with the cost model minimizes the spending for computational services.

1. Introduction

The cloud computing paradigm has been established as a promising model for using computational resources and software resources on-demand [1–3]. As one of its advantages, it shifts complex IT resource management tasks typically performed by a company that is in need of computational resources to service providers. It also increases the flexibility of companies to adapt their IT resources to changes in demand [4,5]. Assuming a cost-based pricing of cloud services, it can reduce infrastructure investment cost and cost for organizing IT resources. The capital expenditure (CAPEX) and operational expenditure (OPEX) are lower due to the economies of scale of the cloud providers' infrastructure.

A potential scenario for the foreseeable future is that of a hybrid cloud environment. Hybrid clouds refer to a composition of a private cloud (i.e., data center managed through cloud technology and owned by the company in need of computational resources) and at least one public cloud (i.e., resources owned by a cloud provider) [6]. Most enterprises, except for those with a very small or stable demand for computing resources, will employ a mix of in-house and outsourced computing resources [7]. The reason for keeping a private cloud, despite the economies of scale of public clouds, is the increased security risks in the public cloud and possibility of serving the base demand at lower cost [8]. Therefore, with the ability of a seamless migration of processing load between a private cloud and public clouds, the hybrid cloud is ideal for handling variable demand [9].

To truly fulfill the promise of cloud computing (i.e., its benefit from economies of scale), technology is needed to federate dis-

* Corresponding author. Tel.: +82 7076786676; fax: +82 15016415384.
E-mail addresses: jorn.altmann@acm.org (J. Altmann), mmkashef@temep.snu.ac.kr (M.M. Kashef).

http://dx.doi.org/10.1016/j.future.2014.08.014
0167-739X/

parate public clouds at the infrastructure level (IaaS). Only through federation and interoperability at this level, small cloud providers (i.e., providers of limited scale) can take advantage of their aggregated capabilities and provide a seemingly infinite service computing utility that can compete with the bigger cloud providers in the market. We refer to the infrastructure that supports this paradigm as federated clouds [10]. By definition, a federated cloud is a cloud, in which (competing) cloud providers have reached a cross-site agreement for cooperating regarding the deployment of service components in a way similar to electrical power providers who use capacity from each other to cope with demand variations among their own customers [11].

To take advantages of hybrid federated clouds, companies need to know the overall cost of running their services on this kind of cloud infrastructure [12–15]. A service runs on one or more virtual machines (VMs) and can be moved (with the VMs) between federated clouds. The reports on the cost savings of clouds vary, however. For example, a report by McKinsey states that moving to the cloud would actually cost 144% more than the current data center infrastructure [9,16]. West reports that cloud technology reduced a government agency's cost by 25%–50% [16]. This uncertainty about cost savings and about actual expenditure were also discussed by Kondo et al. [4]. These examples highlight the necessity for an overall cost model.

Apart from the cost model for federated hybrid clouds, the service placements on this kind of cloud infrastructure need to be considered by the company that is in need of computational resources, as each service placement incurs a different cost. Hence, minimizing the overall cost for using federated hybrid clouds requires an optimization model that considers all cost factors and cost values for each of the service placements [17,18]. Whenever the configuration and prices of clouds change over time, the optimization problem has to be solved again [11,14,19].

Summarizing this discussion, our research can be described with two research questions: (1) What are the cost factors specific to federated hybrid clouds and how can a comprehensive cost model for federated hybrid clouds be defined? (2) If such a cost model for federated hybrid clouds exists, how can it be used for a service placement optimization algorithm?

In particular, we intend to construct an overall cost model that can be used by enterprises to decide where to place their services on a federated hybrid cloud such that it minimizes cost. This overall cost model, which is based on the initial and less comprehensive version of the cost model given in [20], comprises all cost factors and cost functions, which are necessary to estimate the precise cost of running cloud services on an in-house cloud-enabled data center (private cloud) and on federated public clouds.

To answer the research questions, we conduct the following steps: First, we perform a systematic literature review of papers on cost factors and cost models in cloud computing. Second, we identify the gaps in the current research on federated hybrid cloud cost modeling. Based on this result, we design a comprehensive overall cost model for federated hybrid clouds. In the next step, we propose an optimization algorithm for service placement on federated hybrid clouds, using this cost model. Finally, we apply the proposed cost model and the service placement optimization algorithm to a case study to demonstrate its workings.

The contributions of this paper are threefold: First, it comprises a comprehensive cost model that cannot only be used for cost calculation of federated hybrid clouds (as in this paper) but also to guide and compare investment decisions in private clouds. Second, the cost model considers cost for the migration between clouds which includes the deployment cost and the transmission of data between public clouds (and not just between private clouds and public clouds). Third, a brute-force service placement algorithm that considers the cost of service placements.

The remainder of the paper is organized as follows. The next section gives an overview about cost factors and cost models for cloud computing, which have been proposed in literature. Section 3 introduces our cost model for federated hybrid clouds. In Section 4, our cost-model-based service placement optimization algorithm is presented. Furthermore, an application of our cost model as part of the service placement algorithm to a cloud computing scenario is shown. Section 5 concludes the paper with a brief discussion.

2. Background

2.1. Cloud computing, hybrid clouds, and federated clouds

Although there are many definitions of cloud computing, the NIST definition seems to have captured the commonly agreed cloud computing aspects that are mentioned in most of the academic papers [5]. The NIST definition states that cloud computing is "A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [5]. For the purpose of our paper, the definition given by NIST suffices.

Furthermore, we assume that a cloud provider can own several clouds [3]. Each of these clouds is assumed to be at different geographical locations (e.g., different regions of Amazon AWS are considered to be different clouds) and could use the same cloud standard. Besides, from an economic perspective, there is no difference between a cloud provider owning the data center and a cloud provider that rents hardware from a data center provider [3]. The only difference from the perspective of a hardware-renting cloud provider is that this provider outsources the maintenance of the data center hardware. The cloud provider's profit comes from the difference in selling cloud services on demand and the fixed renting cost paid to the data center provider. The data center provider benefits from a 100% utilization of his hardware services and does not need expertise on cloud computing.

The research of this paper is related to two cloud properties, namely, hybrid clouds and federated clouds. The definition of hybrid clouds used in this paper follows the one of Metzler and Taylor as well as Van den Bossche et al. [18,21], in which organizations use public clouds to cover their demand for computational resources in excess of the capacity of their private cloud.

With respect to federated clouds, technology is needed to combine disparate public clouds, including those owned by different organizations. Only through federation (including its interoperability requirement) can a single cloud provider take advantage of the aggregated capabilities to provide a seemingly infinite service computing utility. We refer to this category of clouds as federated clouds [10]. In detail, federated clouds comprise clouds of (competing) cloud providers, who have reached a cross-site agreement for cooperating regarding the deployment of service components (e.g., through a marketplace of standardized goods [22,23]). The concept is similar to electrical power providers, who use capacity from each other to cope with demand variations among their own customers [11].

Fig. 1 illustrates the above definition of cloud federation by showing an example of a cloud customer (company) that is in need of computational resources. This cloud customer runs its private cloud (i.e., its own data center with cloud technology) to host its security-critical services. Beside the private cloud, the company uses two different clouds (i.e., public cloud 1 and public cloud 2) for services that are needed at times of peak demand. The two public clouds offer the same cloud interfaces to each other and the cloud customer, following the cloud federation agreement between the two public clouds.