Future Generation Computer Systems 78 (2018) 102–114

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Enterprise applications cloud rightsizing through a joint benchmarking and optimization approach

Athanasia Evangelinou [a,*], Michele Ciavotta [b], Danilo Ardagna [b], Aliki Kopaneli [a], George Kousiouris [a], Theodora Varvarigou [a]

[a] *National Technical University of Athens, Athens, Greece*
[b] *Politecnico di Milano, Milan, Italy*

HIGHLIGHTS

- A methodology and tools that support the design and migration of applications to Cloud.
- The performance advertised by cloud providers is to be used carefully.
- The proposed benchmark procedure for migrated Cloud applications leads to reduced costs.

ARTICLE INFO

Article history:
Received 7 April 2016
Received in revised form
22 September 2016
Accepted 2 November 2016
Available online 11 November 2016

Keywords:
Benchmarking
Cloud applications
QoS
Model driven design
Performance prediction
Design space exploration

ABSTRACT

Migrating an application to the cloud environment requires non-functional properties consideration such as cost, performance and Quality of Service (QoS). Given the variety and the plethora of cloud offerings in addition with the consumption-based pricing models currently available in the cloud market, it is extremely complex to find the optimal deployment that fits the application requirements and provides the best QoS and cost trade-offs. In many cases the performance of these service offerings may vary depending on the congestion level, provider policies and how the application types that are intended to be executed upon them use the computing resources. A key challenge for customers before moving to Cloud is to know application behavior on cloud platforms in order to select the best-suited environment to host their application components in terms of performance and cost. In this paper, we propose a combined methodology and a set of tools that support the design and migration of enterprise applications to Cloud. Our tool chain includes: (i) the performance assessment of cloud services based on cloud benchmark results, (ii) a profiler/classifier mechanism that identifies the computing footprint of an arbitrary application and provides the best matching with a cloud service solution in terms of performance and cost, (iii) and a design space exploration tool, which is effective in identifying the deployment of minimum costs taking into account workload changes and providing QoS guarantees.

1. Introduction

Cloud computing is a disruptive phenomenon in ICT world, which has rapidly entered mainstream consciousness and gained a significant attention of various communities like researchers, business and government organization. It is obvious, in fact, that everyday services like Dropbox, Netflix or Instagram owe part of their success to the benefits of cloud such as the infinite number of resources, the ability to dynamically adapt (scale-up or scale down) accordingly to usage behaviors and the pay-as-you-go economical model.

With a rapidly increasing number of companies entering the cloud market and offering heterogeneous and constantly evolving technologies [1] the process of software design and implementation experienced a deep change. On the one hand, the cloud has meant for developers providing advanced cloud-based tools and abstractions for development, collaboration and deployment. Dynamic systems capable to react to workload fluctuations by adapting themselves in order to keep the performance unchanged can easily built and run delegating to the cloud provider the intensive tasks of infrastructure management and maintenance. On the other hand, performance unpredictability and vendor lock-in are just some of the issues that developers have to face and prove

* Corresponding author.
E-mail address: aevang@mail.ntua.gr (A. Evangelinou).

http://dx.doi.org/10.1016/j.future.2016.11.002
0167-739X/

significant barriers to widespread cloud adoption. In fact, the increasing size and complexity of software systems, combined with the wide range of services and prices available in the market, puts the designer before the necessity to evaluate a combinatorially growing number of design alternatives [2] with the goal of finding a minimum-cost configuration that suits the application Quality of Service (QoS) requirements.

To carry out such a task, the system designer should consider a large number of alternatives and should be able to evaluate costs (that often depends on the application dynamics) and performance for each of them. This can be very challenging, even infeasible if performed manually, since the number of solutions may become extremely large depending on the number of possible providers and available technology stacks. What is more, in many cases the performance of these service offerings may vary depending on the application types that are intended to be executed upon them and their characteristics in terms of resource usage. Intensive RAM, CPU, disk or GPU usage may be toggled between applications that have different goals, such as scientific components, database instances or front ends.

A key challenge for customers before moving to Cloud, is to know application behavior on cloud platforms, in order to select the best-suited environment to host their application components in terms of performance and cost. In many cases the application owner might not be aware of how the migrated application component uses the available computing resources; for many application components the runtime behavior and usage of resources may not be known, mistakenly considered or altered due to structural changes during the software migration. On the other side, cloud providers might also be interested in knowing the application types hosted within their infrastructures to avoid interference effects of concurrently running VMs, which significantly degrade applications performance. Evaluating a specific and arbitrary application component over the entire range of offerings becomes a daunting task, especially when the deployment of the former may be subject to provider-specific or component-specific actions or code in each case. However, finding a more abstracted and common way for identifying both application profile and performance aspects of cloud environments may significantly reduce the effort needed in this task.

Finally, the performance characteristics of a cloud may change over time dramatically, depending on the congestion level, policies implemented by the cloud provider, and competition among running applications. Assessing the performance of an application in the cloud is a complex process that requires unbiased data and specialized models often implying skills that go beyond those commonly exhibited by software engineers. This situation calls for analytical techniques, models, application profiling and benchmarks that simplify the process of performance evaluation at design-time in order to support the user in the decision making process.

This work aims at proposing a methodology and a tool chain that support the migration and pricing scheme of enterprise applications to the cloud. Our tool chain exploits cloud benchmarking results and includes a profiler/classifier, which identifies the computational nature of a software component in a black box manner and a design space exploration tool, which is able to identify the cloud configuration of minimum cost fulfilling QOS constraints taking into account also daily workload patterns.

The remainder of the paper is organized as follows. In Section 2 an overview of the design methodology and the implemented tools are presented, while in Section 3 the benchmarking results acquisition is described. Section 4 introduces the profiling and classification tools, while Section 5 is devoted to the design time exploration tool. In Section 5 the experimental results achieved on a case study by combining and integrating the implemented tool chain are presented. In Section 7 related work in the respective fields is described. Conclusions are finally drawn in Section 8.

2. Overview of the design methodology and tools

The combined methodology of the tool chain appears in Fig. 1. Our approach along with the supporting tools, as an "all in one" solution allows the applications to exploit the offerings of the cloud providers in terms of performance and cost, taking into consideration their extracted computational behavior via profiling. The implemented methodology includes three phases: (i) Benchmarking (ii) Profiling and Classification and (iii) Assessment and Optimization phase.

Once the benchmark application types have been selected, the respective benchmark tests are executed automatically through the *Benchmarking Suite* on the candidate target cloud infrastructures, and their results are used to populate the *Raw Data DB*. Next, the application VM along with the benchmark application types are profiled and the computational profile of each is identified. When the profiling process is completed, the obtained computational profiles are used as an input to the *Classification Tool*. The latter determines the optimal cloud service solution for the given application in terms of performance and cost restricting the number of alternatives to be tested in the Optimization phase. Then, the stored results from the *Raw Data DB* along with the set of candidate providers and instance types obtained from the classification process, are imported within *Space4Cloud* resource DB. In the last step, the imported results are exploited during the *SPACE4Cloud* candidate solution performance assessment to evaluate how the performance metrics of the target application changes by varying the type and size of the hosting resources for subset of cloud providers that are considered as a target of the final deployment.

3. Cloud benchmarking

When considering the migration of existing applications to the cloud, it is critical to examine both the diversity of cloud providers and the varying performance issues of cloud services. Since there is an increasing number of providers offering cloud infrastructures and services a fair evaluation of such cloud systems is needed. System architects and developers have to tackle with this variety of services and trade-offs. Moreover, in some cases cloud providers offer their own metrics for evaluating and guaranteeing cloud QoS.

Hence, in order to measure performance aspects and select the cloud services that fit best to the application to be migrated, an abstracted process is implemented by using suitable tests and tools, namely benchmarking. The first step of this process is to define a set of performance stereotypes based on different application categories. The main goal of these stereotypes is to extract a number of performance characteristics of the provider that are necessary for meeting QoS requirements of the migrated cloud applications. The source of these characteristics are common application types that correspond to various popular applications and have been linked to respective benchmarks that can be used to indicate a specific offering ability to solve real-life computational problems. Thus, tests have been identified with specific workload patterns that can be mapped to concrete real world applications. Benefits of such a categorization include the ability to abstract offering performance capabilities on an application description level, thus being easily ranked according to user interests for a specific category.

Concerning the characterization of a service ability from a performance point of view, we use the *Benchmarking Suite* for benchmarking cloud platforms in order to extract performance-related data [3]. The set of the application types are reported in Table 1. The specific benchmarking tools provide a large number of application-level benchmarks which incorporate and reflect characteristics of CPU, I/O, network and data-intensive real-world applications. Therefore, these tools can stress the performance