

Chapter 5

Resource Allocation for Cloud Computing: A Semantic Approach

Jorge Ejarque

Barcelona Supercomputing Center, Spain

Javier Álvarez

Barcelona Supercomputing Center, Spain

Raül Sirvent

Barcelona Supercomputing Center, Spain

Rosa M. Badia

Barcelona Supercomputing Center, Spain & Spanish National Research Council, Spain

ABSTRACT

Cloud computing has emerged as a distributed computing paradigm where resources are requested on demand and in a very dynamic fashion and paying only for what you consume. This new paradigm created an ecosystem where several providers offer heterogeneous computing resources to satisfy the customers' computing demand. So, the allocation and adaptation of this demand to the correct resources is a key issue in this ecosystem, because it can produce a mutual benefit for the customers and providers. However, with the wide variety of customers and providers, this allocation is not an easy task. This chapter presents a toolkit that implements a methodology for improving the resource allocation between different Cloud providers. The Semantically Enhanced Resource Allocator (SERA) toolkit introduces the semantic web and multi-agent technologies for facilitating the interoperability between the users and different resource providers. Semantic web technologies provide the required semantic interoperability between the different providers' vocabularies; meanwhile a platform of configurable agents provides an adaptable and autonomous way of allocating and managing execution requests and resources according to the customers and providers rules.

DOI: 10.4018/978-1-4666-0098-0.ch005

INTRODUCTION

In the cloud ecosystem (Vaquero, Roderio-Merino, Caceres, & Linder, 2008), we can find different providers which offer computing Infrastructure as a Service (IaaS) on demand. This group of providers and their resources has been called *Public Cloud*. With the evolution of virtualization technologies, enterprises start to organize their big datacenters as *Private Clouds* dedicated for deploying services and executing computational tasks which are required by their business. This virtualized organization of datacenters facilitates the management and configuration of the computing infrastructure, automating repetitive processes as well as allowing the outsourcing of computation to Public Clouds in peaks of computing demand. However, the outsourcing of computation is not an easy task: it has to coordinate different actions and raises several problems. The different Cloud Providers offer heterogeneous resources described according to their own metrics or properties which are different in other providers. Those descriptions must be understood by the customers and compared to their requirements. Once the descriptions have been understood, the providers' resources must be allocated to customers' tasks taking into account not only the technical requirements but also the customers and providers rules. After the resource allocation, the customers' tasks must be executed on the selected resource and, finally, the execution must be monitored for detecting unexpected events adapting the execution for solving them.

This chapter presents the Semantically Enhanced Resource Allocator (SERA), a toolkit for allocating customers' tasks to resources from different resource providers benefitting from the semantic web (Berners-Lee, Hendler, & Lassila, 2001) and multi-agent (Ferber, 1999) technologies. On one hand, semantic technologies are used to achieve semantic interoperability between resource descriptions from different providers

and different customers' requests. On the other hand, agents increase the autonomy and self-management of a system. Agents are proactive, so, they can take decisions by themselves according to their beliefs, desires and intentions and trigger actions in the system by their own initiative. For this reason, agents are suitable for coordinating the allocation and execution of tasks in these clouds environments, detecting execution problems and reacting to them by triggering the most appropriate actions depending on the system status. They are also capable to communicate each other implementing negotiation protocols which are very useful to reach agreements with other agents. Benefiting from this capability, the resource allocation in our methodology is agreed by means of a negotiation between the customer and provider agents taking into account the customers and providers rules.

The capabilities presented above enable the SERA toolkit to support different customers' requests (such as batch jobs and service deployments) and to interoperate with different computing platforms (such as Grids and Clouds) giving a general resource allocation solution. For this reason, some SERA parts (such as components, agents and ontology classes) are presented in general. However, in the different chapter sections we have tried to give details about how these concepts specifically apply to the Cloud Computing environment. The chapter is organized as follows: first section presents the SERA toolkit architecture, focusing in the most important concepts: the ontology for resource allocation in clouds, the automatic annotation of resource descriptions from the different providers' schemes and the multi-agent based semantic resource allocation process. Following, we provide installation and deployment guidelines as well as different examples on how to use the SERA toolkit. The chapter is finalized drawing conclusions about the lessons learned and presenting future research directions.

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/resource-allocation-cloud-computing/62366

Related Content

SAT and Planning: An Overview

Carlos Camarão, Mateus Galvão and Newton Vieira (2010). *Intelligent Systems in Operations: Methods, Models and Applications in the Supply Chain* (pp. 27-40).

www.irma-international.org/chapter/sat-planning-overview/42653

Cloud Computing: Security Concerns and Issues

Shantanu Pal (2013). *Cloud Computing Service and Deployment Models: Layers and Management* (pp. 191-207).

www.irma-international.org/chapter/cloud-computing-security-concerns-issues/70141

The Environmental Awareness Influence Urban Female's Purchasing Intention of Organic Agricultural Products

Jing-Jing Lai, Nai-Yuan Pai and Wen-Cheng Wang (2020). *International Journal of Information Systems in the Service Sector* (pp. 1-18).

www.irma-international.org/article/the-environmental-awareness-influence-urban-females-purchasing-intention-of-organic-agricultural-products/252202

Effect of Mediating Factors on Customer Defection with Special Reference to Banking Industry in India

Sunita Dwivedi and Bharti Wadhwa (2016). *Handbook of Research on Promotional Strategies and Consumer Influence in the Service Sector* (pp. 138-153).

www.irma-international.org/chapter/effect-of-mediating-factors-on-customer-defection-with-special-reference-to-banking-industry-in-india/149723

Integrated Fog and Cloud Computing Issues and Challenges

Shivom Sharma and Mohammad Sajid (2021). *International Journal of Cloud Applications and Computing* (pp. 174-193).

www.irma-international.org/article/integrated-fog-and-cloud-computing-issues-and-challenges/288780