Agent technology based resources management in Cloud Computing

Taha Chaabouni  
MIRACL Lab, FSEG  
University of Sfax  
Sfax, Tunisia  
chaabounitaha@yahoo.fr

Hamdi Kchaou  
LOGIQ Research Group  
University of Sfax  
Sfax, Tunisia  
kchaou_hamdi@yahoo.fr

Maher Khemakhem  
MIRACL Lab, FSEG  
University of Sfax  
Sfax, Tunisia  
maher.khemakhem@fsegs.rnu.tn

Abstract—This paper proposes a model based on agent technology for the optimization of resources management in Cloud Computing. We focus indeed on the providers side in order to provide them a tool which is able to optimize automatically their resources management in a customized manner since we observed the lack of such a tool and this to the best of our knowledge.

Keywords—cloud computing; data center; multi agent system; system management

I. INTRODUCTION

Cloud computing is relatively a new concept. It combines different existing technologies like: Service Oriented Architecture, grid and distributed computing, virtualization... It represents a model used to deliver IT capabilities. In cloud computing, everything (any given resource) is provided as a service. Consequently, the service provider reacts according to the user (client or customer) request by providing him all resources that satisfy his requirement. We distinguish between three main services: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). In SaaS, a full ready application is available to the customer. In PaaS, the user has the development tools to create his own application. Finally in IaaS, the infrastructure provider prepares the hardware part as a service. This computing infrastructure is offered to the client as Virtual Machine (VM). VM is a software implemented machine which runs on physical machine. VM works like a real machine and can run different operating systems and applications since it has the same behavior. One physical machine can execute in the same time several different VM. Pay as you go is the most important advantage of cloud computing. It has a direct effect in the cloud development. In the actual economic situation (world crisis), many companies attempt to reduce their financial expenses by using cloud services. So, why to buy a computing infrastructure when it can be rent. Different situations have influenced the development and the adoption of cloud computing.

The U.S. National Institute of Standards and Technology (NIST) proposes [1] the following definition: “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”.

This definition seems to have captured the commonly agreed aspects of cloud computing, was adopted by the most researchers. It is important to agree on one definition to limit the scope of research and emphasize the potential business benefits.

Cloud computing is still in developing stage. Many challenges are not resolved yet especially optimization and scheduling problems. In fact, a cloud provider deals with thousands of physical machines and up to millions of VM. The challenge for him is to satisfy the request of the client with a quality of service. It is important to create an efficient strategy to manage this entire infrastructure. In this work we present a management strategy (model) of cloud computing based on agent technology which attempts to ease the control of cloud computing. This work is composed of two sections. In the first one, we summarize some researches related to this work. The second section is dedicated to present our proposed management system for cloud computing.

II. RELATED WORK

A. Resource Management

Optimization in cloud computing represents one of the most active research areas. The motivation was the huge number of the resources used in the system of cloud. Generally in cloud, there are two principal actors: the service provider and the service consumer. Depending on the point of view, there are different strategies to use cloud resources: the service provider want to reduce the cost of running the system and the consumer wish to reduce the cost of using the cloud infrastructure with the minimum execution time. The service provider must consider client’s constraint and in the same time reduce his running cost. Any failure of this challenge can lead to serious loss.

Data center [2], operating under the cloud computing model, can host different types of running applications starting from those that work for few seconds (web application like e-commerce) to those that need a longer periods of time (simulation or large data processing). The management of this
various types of applications in one data center creates a challenge between on demand resource provisioning and allocation in response to time varying workloads.

Monitoring of cloud is one of the important tasks in the cloud management. It supplies the service provider with critical information like the workload, the performance and the Quality of Service (QoS). This information is important to keep a track of the hardware state to prevent a future failure that lead to a stop of service. The most important activities, in cloud computing, which need monitoring, are [3]: resource planning and management, data center management, troubleshooting, performance and security management. So, a monitoring and data analysis system must be implemented to have a good management strategy.

B. Agent Technology in Cloud Computing

Agent technology is used to benefit of the intelligent behaviors, communications and negotiation capabilities. In cloud computing we have two actors: the client and the service provider. Different works have presented ideas to create a negotiation system between these two actors to gain the best cost for the client. We have identified out different specifications for the agent: like finding out the best offer or discovering a new cloud computing provider...

The work [4] presents an agent-based testbed system to bolster cloud resource discovery and service level agreement negotiation. There are different jobs for the agents in this system. Actually, every action is achieved by a specific agent. There is an agent responsible for finding out all cloud computing systems that can satisfy the client request. Another agent is dedicated to negotiate a given phase or step. Depending on the client requirements, this agent will choose the best offer presented by cloud providers. It can even split the client work to different parts on different cloud systems to obtain a better result.

The work [5] proposed a complete system based on agent technology for provisioning and management. An application was developed to help the client in the different phases starting by the cloud provisioning (selecting the cloud resource), then the cloud configuration and finally cloud monitoring.

To the best of our knowledge there is no previous work which has used agent technology to optimize the resources management for the service provider.

III. PROPOSED WORK

A. Problem description

The management of cloud computing in IaaS level is a difficult task for the service provider. In fact, it can’t be achieved manually. In a simple cloud computing system which is composed of one data center, we can easily find out thousands of physical resources that are virtualized in millions of virtual resources. So, the manager needs an automatic solution to supervise the system in order to minimize the human intervention. On the other hand, this system must allow some leeway to the manager so he can set his own rules (customization) to supervise the system or adapt them to a new situation. Multi-agent system is a method that can be adopted as a solution to this problem. In fact, this solution offers an automatic management of the cloud system with a quick adaptation to the new situation. So, we attempt to deploy a multi agents system to manage the cloud computing infrastructure. These agents will be charged to monitor the different resources. This control deployed locally in the same data center or globally in the cloud computing system. To control the agent, a management interface is important to simplify the task of the cloud computing manager. Our work is composed of three essential phases. In the first part we define and describe the agent system (interaction, communication, type of agent...), then we create and define the agent and finally we create the management interface and the list of rules to be taken into account by the agent who will be charged in the resources management. We look to obtain a fully automatic system in which the client can connect to a web site, input his requirements and loads his work. At this moment, the agents system will satisfy this request in an optimal manner according to the provider rules already customized.

B. Agent system description

The Multi-agent system is composed of two types of agents: Supervisor Agent and Data Center Agent. We have one instance of the agent Supervisor and for each data center an agent named Data Center Agent is created. The proposed system is based on communications between agents, and it aims to minimize the response time of client’s demands. To satisfy this constraint, all agents used in the system must be reactive. So, we use here a centralized architecture of agents which uses simple exchange of messages between the Supervisor Agent and Data Center agents.

![Figure 1: Integration of agent in cloud computing system](image)

The Supervisor agent receives the client’s demands and treats them one by one; the first come is the first served. The Supervisor agent sends a message to the Data Center agents and asks them about their available (idle) resources. Then,
every Data Center agent should reply to the Supervisor agent. This reply includes the corresponding status (idle or not) and how many resources this Data Center can offer. After receiving responses from all Data Center agents, the Supervisor agent selects the best choice based on two rules.

![Diagram of Communication between Agents](image)

**Figure 2. Communication between Agents**

First rule: if there are not enough available (idle) resources, the Supervisor Agent will wait for a message from the first Data Center Agent that will indicate that he has the free resources requested.

Second rule: if there is more than one Data Center which has the required resources, the Supervisor agent selects the nearest one in term of distant.

Remarque: every Data Center Agent that has a finished work sends a message to the Supervisor agent in which is mentioned his new available resources.

**IV. CONCLUSION**

In this work, an under development system for cloud management based on agent technology was presented. This system focuses on the service provider management resources. It attempts to control automatically the cloud computing resources. In this case, we don’t need a human intervention. The design phase and the definition of the system features were presented in this paper. As a future work, we will complete the implementation of the system and try to integrate it in cloud computing system like OpenNebula or Eucalyptus.

**REFERENCES**


