

A QoS-oriented Inter-Cloud Federation Framework

Maria Salama
British University in Egypt
Cairo, Egypt
Email: maria.salama@bue.edu.eg

Ahmed Shawish
Ain Shams University
Cairo, Egypt
Email: ahmed.gawish@fcis.asu.edu.eg

Abstract—Cloud federation allows individual cloud providers dynamically collaborate to offer services to their end-users with the Quality of Service (QoS) targets agreed in the Service Level Agreements (SLA). However, the current federated cloud models are not QoS-oriented or SLA-aware. This paper proposes a QoS-oriented federated cloud computing framework where multiple independent cloud providers can cooperate seamlessly to provide scalable QoS-assured services and discusses a high level architecture of the federation components. The distinct features of the proposed federation framework is its QoS-orientation that can trigger the on-demand resource provisioning across multiple providers, hence helping to maximize QoS targets and resources usage, eliminate SLA violations and enhance SLA formalization.

Keywords—cloud federation; quality of service

I. INTRODUCTION

Cloud Providers are facing significant challenges with QoS targets claimed in SLAs agreed with end-users; especially in the context of delivering low-cost high-performance services with the unpredictable end-users workloads that might reach its peak at any time. Such circumstances will report service failure in a frequent manner, SLA violation and consequent results of financial penalties and reputation loss. Cloud federation facilitates dynamic expansion of services across multiple providers to achieve QoS targets under variable workloads and computing resources. The current federated cloud computing models are not QoS-oriented; they are not tailored to achieve QoS targets as primary goal and guaranteeing SLAs. In this paper, we propose a QoS-oriented federated cloud computing framework where multiple independent cloud providers can cooperate seamlessly to acquire more resources in peak times to fulfill their QoS targets and pre-define SLAs new QoS levels in terms of their scaled resources after performing federation agreements. We discuss a high level architecture of the federation components. The framework includes formalization of federation agreements, workload prediction and QoS continuous evaluation, besides resources monitoring and allocation. The framework components offer QoS-oriented capabilities to address dynamic resources management, aiming to dramatically improve the effective usage of resources. This will provide not only enhanced better degrees of scalability and reputation, but also higher QoS targets based on the new scale of resources.

II. BACKGROUND

Despite the fact that the federated cloud computing concept is recently emerging, the proposed federation frameworks and models were not QoS-oriented; they are not tailored to achieve

QoS targets as primary goal and guaranteeing SLAs. Hence, such models cannot be a real backup for providers in cases of peak demand or service failure. RESERVOIR model for open federated cloud computing [1] only considers management of the server virtualization. InterCloud [2], a utility-oriented federation of cloud computing environments for scaling of application services, facilitates scalable provisioning of application services under variable workload, resource and network conditions. Another business-oriented federation model for real-time applications [3] was taking into consideration only critical requirements for real-time applications. This necessitates building federation mechanisms for provisioning resources across providers to deliver on demand, cost-effective and QoS-oriented services.

III. PROBLEM DESCRIPTION

Cloud computing is an on-demand service delivery based on a guaranteed QoS with the client, as agreed in the SLA between the client and the provider. Meanwhile, unpredictable client requests might reach its peak in certain times. To provide the service with the guaranteed QoS, especially in peak times, cloud providers find great challenge to allocate resources based on a widely varying, unpredictable client requests and resources of the provider could not enough to maintain the same level of QoS agreed with the customers. For the cloud service provider, any violation in the QoS agreement may result in financial penalties charges, as well as loss of customers and reputation in case of frequent failures. Cloud federation allows individual cloud providers to engage in an agreement with other cloud providers to enable dynamic resources allocation. Based on the imposed situation, the provider would perform federations with others; so that they act as his backup at peak times. Federations need to be more SLA-aware. Since the QoS attributes change constantly in a dynamic environment, the federation framework should be dynamically QoS-oriented.

IV. VISION AND REPRESENTATION

To meet aforementioned requirements of auto-scaling cloud providers resources, an inter-cloud federation framework should focus on design, development, and implementation of software systems and policies for federation of clouds based on the QoS targets that are agreed in SLAs with end-users. These Federation Agreements would help them in providing better service with higher QoS targets based on the new scale of resources gained from the federation. The key elements for enabling cloud federation would be Cloud Coordinators (CC) and Federation Coordinator (FC). The resources provisioning within the federated clouds would be will be driven by QoS-oriented principles for efficient resource allocation depending

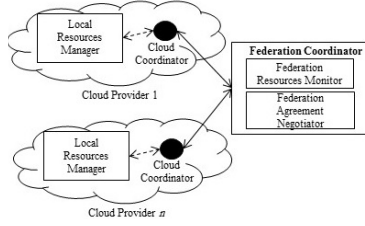


Fig. 1. Cloud Providers Federation

on QoS targets and workload demand patterns. To maximize the providers benefit out the Federation Agreement, an SLA formalization helper will let providers offer better QoS levels gained from scaled federated resources.

The approach for realization of this research vision consists of investigation, design, and development of the following:

- Architectural framework and principles for QoS-oriented clouds and their federation.
- Formalization of Federation Agreements between providers.
- Formalization of SLAs between providers and end-users based on the better QoS levels gained from scaled federated resources.
- An FC responsible for mediating between Cloud Coordinators of different providers and acting for enabling resources provision across multiple providers.
- A CC for managing agreements services (Federation Agreements and SLAs), resources allocation and scheduling driven by QoS targets.
- A software platform implementing CC, and FC.

Figure 1 shows the federation framework. The FC is responsible for mediating between CCs of different providers. FC allows the CC to locate resources from other providers at the demand time. The FC acts as an information registry that stores the current resources usage of the federated clouds. Furthermore, it provides match-making services that map CCs requests to providers having available resources. The Federation Agreement specifies the details of the resources a provider would offer to the federation. Every provider within the federated environment needs to instantiate a connection between his CC and the Federation Agreement Negotiator that can dynamically establish service contracts.

The CC is responsible for managing the providers resources and its interaction with the federation. The CC identifies a suitable cloud service provider through the FC and negotiates for an allocation of resources to meet his QoS needs. The high level architecture of CC is shown in Figure 2 and its components are discussed below.

The QoS Evaluator is measuring the quality parameters performed at run time. This continuous evaluation early detects any possible violation for the QoS targets and alerts the Federation Negotiator. The Workload Predictor determines if a peak of jobs is expected. Since workloads are varying, it is essential to forecast peak workloads.

The Resources Monitor maintains the status of resources by periodically checking their availability with the Resources Allocator and discovering new available resources. It alerts the Federation Negotiator if there is a possibility that current resources will be unable to satisfy the SLA requirements or if the availability of local resources is reaching the safe limit.

The Federation Negotiator bargains with the FC for resources allocation that meets QoS needs, once alerts are received from

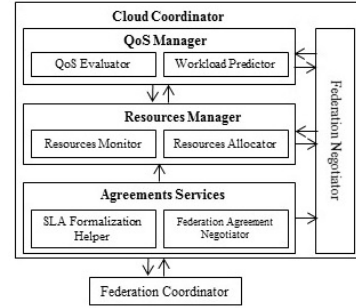


Fig. 2. Cloud Coordinator (CC)

either the Workload Predictor or the Resources Monitor. In this case, the Federation Negotiator request that encapsulates the QoS requirements is submitted to the CC. The matching procedure considers the QoS targets and other providers resources utilization. In case the match occurs, the quote is forwarded to the Service Negotiator. Following that, the Resources Allocator deploys the requests to the resources available from the provider suggested. The Federation Negotiator also keeps track of available resources in off-peak times to be offered to the Federation.

The Resources Allocator is responsible for allocating available resources, either local or accessed from the federation based on QoS targets and SLA requirements. On receiving a client request, the Resources Allocator distributes the workload between local and federated resources in case extra resources are given from the federation; and updates the Resources Monitor with the currently consumed resources.

The Agreements Services component stores the service terms and conditions that are being supported by the Cloud to the federation. The Federation Agreement Negotiator establish service contracts with the federation environment. The SLA Formalization Helper ensures that the provider would offer better SLAs to his end-users based on the better QoS levels gained from scaled federated resources.

V. CONCLUSION

In this paper, we propose the idea of a QoS-oriented inter-cloud federation, where providers dynamically act as backup for each other in peak times. The resources provisioning within the federated clouds is driven by QoS-oriented principles for efficient resource allocation. The workload prediction and QoS evaluation will protect the provider from any possible SLA violation. The proposed framework facilitates the federation management and provides a QoS-oriented and highly dynamic federated environment. This will provide not only enhanced better scalability and reputation, but also higher QoS targets based on the new scale of resources gained from the federation.

REFERENCES

- [1] B. Rochwerger, D. Breitgand, A. Epstein, D. Hadas, I. Loy, K. Nagin, J. Tordsson, C. Ragusa, M. Villari, S. Clayman, E. Levy, A. Maraschini, P. Massonet, H. Muoz, and G. Tofetti, Reservoir - When One Cloud Is Not Enough, in *Journal of Computer*, vol. 4, issue 3, March 2011.
- [2] R. Buyya, R. Ranjan, R. Calheiros, InterCloud: Utility-oriented Federation of Cloud Computing Environments for Scaling of Application Services, in *Proc. 10th Int. Conf. Algorithms and Architectures for Parallel Processing*, May 2010, Busan, Korea.
- [3] X. Yang, B. Nasser, M. Surridge, S. Middleton, A business-oriented Cloud federation model for real-time applications, in *Future Generation Comp. Systems*, vol. 28, issue 8, 2012.