LOG8415
Advanced Concepts of Cloud Computing

Foutse Khomh
S. Amirhossein Abtahizadeh
Département Génie Informatique et Génie Logiciel
École Polytechnique de Montréal, Québec, Canada
foutse.khomh[at]polymtl.ca
a.abtahizadeh[at]polymtl.ca

1 Identification

Student's name: Chun-An Bau

Date of the reading note: Sep 21, 2021

Author(s): Yahya Al-Dhuraibi, Fawaz Paraiso, Nabil Djarallah, Philippe Merle

Title of the article: Elasticity in Cloud Computing: State of the Art and Research Challenges

Publication: IEEE Transactions on Services Computing, IEEE, 2018, 11 (2), pp.430-447.

2 Article

Keywords: Elasticity, Cloud Computing, Auto-scaling, Resource provision, Scalability, Containers

Concepts and definitions:

- Elasticity: The ability to modify the quantities of resources during the execution, classified into horizontal and vertical elasticity; the former consists of adding or removing instances of computing resources associated with an application, while the latter consists of increasing or decreasing computing resource characteristics.
- Cloud Computing: Technology for customers to use computing infrastructures such as CPU capability, network, and storage provided by cloud service providers.
- Virtualization: The technology to divide hardware resources into multiple instances and to allow them to execute simultaneously. Hypervisors are required to manage the works in most cases.

- Container: Also referred to as OS-level virtualization, it is more lightweight and resource-efficient than traditional virtualization. The instances on the container that shared the identical OS and guest OSs are required.

Summary: The article focuses on illustrating the elasticity of cloud computing with lots of existing solutions.

The paper explains the definition of elasticity and the classification of different elasticity mechanisms based on the characteristics: configuration, scope, purpose, mode, method, provider, and architecture. Also, the review of elasticity evaluation tools and platforms implemented in diverse works is made.

To improve elasticity in cloud computing, containerization is involved in some approaches. The article covers the introduction about the approaches that use the technology and discuss the benefit or disadvantage of using containerization.

Although the researchers have been studied for the elasticity issues of cloud computing for years and considerable works have been published, the technology still has some challenges to encounter. Therefore, the authors provide some directions at the end of the paper.

Research contributions:

- Propose precise definitions of elasticity as well as related concepts.
- Provide extended classification for elasticity mechanisms according to the configuration, the scope of the: configuration, scope, purpose, mode, method, provider, and architecture.
- Discuss the existing container technologies and their relation to cloud elasticity.

3 Analysis

Quality:

General organization:	Language and style:	Technique:	Bibliography:
☐ Very good;	☐ Very good;	☐ Very good;	■ Very good;
■ Good;	■ Good;	☐ Good;	☐ Good;
☐ Medium;	☐ Medium;	☐ Medium;	☐ Medium;
☐ Bad;	☐ Bad;	☐ Bad;	☐ Bad;
☐ Very bad.	☐ Very bad.	☐ Very bad;	☐ Very bad;
		■ N/A.	

Forces of the message:

- Providing a clear definition of elasticity and novel classification for elasticity mechanisms.
- List plenty of related solutions and summarize their contributions in a plain paragraph. Also, a table to classify them into specific divisions is presented in the paper.
- Some interesting and worth researching directions, which can improve the utilization and user experience, are given.

Weaknesses of the message: Although the paper provides lots of future directions, I can't entirely agree with one of them. Start-up time is the time needed to allocate resources in response to the client's demand for resource adjustment. In the authors' opinion, the cloud service provider should not charge the customers until the adjustment is fully set up. However, in my opinion, the start-up time should be charged without a doubt since the resources are occupied by the customer who made the query and cannot allocate to others. Moreover, this is a generic situation existing in the cloud computing industry. There will be a time lag between the demands and the response from the cloud service providers. No matter the query is launching instances, turning instances down, or adjusting required resources. So, it is not appropriate to put the issue in the future works section of a paper that focuses on discussing the elasticity of cloud computing according to my way of thinking.

Nonetheless, *"The lower the start up time is, the better the elastic solution is. Higher start up time affects the efficiency of elasticity system"* is mentioned at the end of the section. But the sentence doesn't make sense for me, and I cannot even see the relevance between the start-up time and the quality of the solution.

Future directions: The paper also lists the current challenges of cloud computing elasticity. Some of the interesting issues are:

- Interoperability: It is difficult to combine multiple clouds to provide resources since each cloud provider has its own technology and techniques. In order to do so, standardized APIs are required, and the industry needs to agree on standards. Although some approaches have successfully allocated resources from different providers or data centers, they are limited to certain criteria.
- Container-based elasticity: According to the consideration of management and cost, the resources for a container are set with static limits, such as Docker use cgroups to limit the resources consumed by a container. The authors indicate the following directions can be researched in the future.
  - Can we apply the elasticity solutions used in VM on the containers?
  - How to use proactive approaches to anticipate container resource usage and react in advance to scale up/down resources?
- Combined elasticity between VMs and containers: Cloud providers use containers on the top of virtual machines to have many instances arranged across levels of hierarchy. However, resizing container resources is limited by the resources of the virtual machine. The direction is worth researching since achieving elasticity control for VM and containers will allow great flexibility and would be an efficient elasticity solution.

Other important articles:

- A Survey and Taxonomy of Self-Aware and Self-Adaptive Cloud Autoscaling Systems
  ACM Computing Surveys, 51, 3, Article 61 (April 2018), 40 pages
- An elastic controller using Colored Petri Nets in cloud computing environment
  Cluster Computing 23, 1045–1071 (2020).
- A control theoretical view of cloud elasticity: taxonomy, survey and challenges
  Cluster Computing 21, 1735–1764 (2018).
- Horizontal and Vertical Scaling of Container-Based Applications Using Reinforcement Learning
  2019 IEEE 12th International Conference on Cloud Computing (CLOUD), 2019, pp. 329-338