GSRT
GENERAL SECRETARIAT FOR RESEARCH AND TECHNOLOGY

HELLENIC REPUBLIC
UNIVERSITY OF MACEDONIA

A Techno-Economic Assessment of Microservices

Ioannis Papakonstantinou
mai19056@uom.edu.gr

Sarantis Kalafatidis
kalafatidis@uom.edu.gr

Lefteris Mamatas
emamatas@uom.edu.gr

Department of Applied Informatics - University of Macedonia
Thessaloniki, Greece

16th International Conference on Network and Service Management
2-6 November 2020

CNSM 2020

Motivation & Research Challenges

The microservices design paradigm:

- Offering flexibility of cloud computing, scalability, fault-tolerance and resource-allocation benefits
- Introduce virtualization and communication overhead

Service Provider (SPs) target:

- Efficient deployment and operation of microservices, focusing on the maintenance of server resource allocation
- Efficient resource management, leads to profit increases
- Tuning of user Quality of Experience (QoE) and infrastructure monetary cost trade-off

In this work, we highlight:

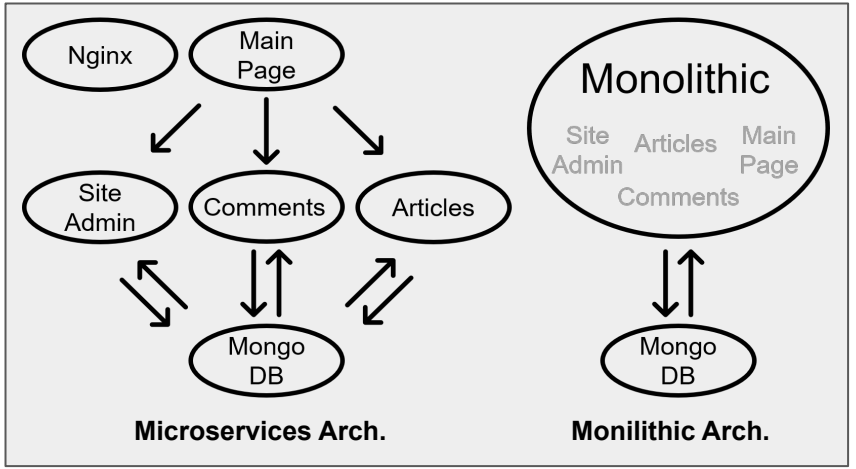
- The technical capabilities and cost-saving impact of microservices in contrast to traditional monolithic applications
- The service performance vs resource allocation trade-off, while elasticity is driven from service quality metrics
- The balance between SP's profit margins and their customer's satisfaction, i.e., reducing the infrastructure cost while keeping the service performance at an acceptable level

Our Approach

We carried out a comparative analysis of two variations of a web application based on both microservices and monolithic architectures.

Methodological choices and assumptions:

- Services are deployed in containers to decouple the impact of MA technology used
- Resource limitations are set to the containers to emulate a distributed deployment
- Users' satisfaction (i.e., QoE) depends on the average Response Time (RT)
- Infrastructure costs are considered from the CPU utilization view-point



The techno-economic analysis that follows:

- Compares experimentally the monolithic with the microservices version of the studied web application, assuming realistic patterns of user visits
- Investigates the potential of alternative application deployments to handle a large number of users with tolerable performance, while reducing significantly the cloud resource expenses

Service Architectures Experimental Analysis

Experimental analysis between different service architectures and elasticity strategies

We investigate CPU consumption and the the service RT of three different cases of the studied web service:

- A Monolithic version that utilizes enough containers to support the given number of users without violating the SLO
- A CPU-based MA deployment that scales the microservices based on a 80% CPU threshold
- A SLO-based MA deployment that attempts to maximize cost savings through scaling the microservices according to the chosen 3 sec SLO value

We emulate 17,500 requests and we observe:

- The MA overhead increases the RT but is not violate the SLO
- An inverse relation between the CPU Cores and the RT
- The SLO approach produces significant cost savings
- The flexibility offered by the MA, since not consume significant resources
- The Monolithic case leading to inefficient resource management while the entire application needs to be scaled up
- There is even more space to trade service performance (i.e., RT) for more efficient resource management

Services		Monolithic	CPU-based MA	SLO-based MA
Monolithic		5.00	0.00	0.00
Microservices	Articles	0.00	0.50	0.10
	Site Admin	0.00	0.10	0.10
	Comments	0.00	0.10	0.10
	Main Page	0.00	0.10	0.10
	Nginx	0.00	1.00	1.00
	Total	5.00	1.80	1.40

CPU Cores required for the 3 application deployments.

Deployment	CPU Cores	Avg. RT	Med. RT	95% RT	Er. %
Monolithic	5	76	70	138	0
CPU-based MA	1.8	648	606	1233	0
SLO-based MA	1.4	1529	1575	1814	0

CPU Cores consumed & statistical data of the measured RT

Cost Analysis

We investigate the infrastructure costs for the three different application deployments.

Based on real cloud prices and user requests values according to the Amazon Web Services and we consider:

- A high workload scenario with over 2 bil. visits per month scaling up our previous results with this number of visits (Table I).
- An additional experiment that emulates 375,000 visits to validate our findings (Table II & Table III)

We observe:

- The cloud resource costs in the case of Monolithic service deployment is around three times higher than the CPU-based MA and SLO-based MA deployments, respectively (Table I)
- The average response time in Monolithic configuration is lower by 194ms, but requires the double amount of resources, in terms of CPU Cores, compared to SLO-based MA configuration (Table II & Table III)

Deployment	Exp. Cores	Pred. Cores	Servers	Cost
Monolithic	5	825	104	$77,688
CPU-based MA	1.8	297	38	$28,386
SLO-based MA	1.4	231	29	$21,663

Table I: Simple Cloud Cost Analysis

Deployment	Containers	Avg.	95% Line
Monolithic	10 Monolithic	1,737	2,611
SLO-based MA	3 Articles - 2 Nginx	1,931	2,309

Table II: Performance Evaluation (Response Time)

Deployment	Cores	Total Cores	Servers	Cost
Monolithic	10	16	2	$1,494
SLO-based MA	5.3	8	1	$747

Table III: Actual Cloud Cost Analysis

Conclusions & Next Steps

We conclude that:

- Although MA introduce communication and containerization overhead, the flexibility introduced exceeds this barrier and offers efficient resource management, leading to significant cost savings for the service provider
- MA can produce significant cost savings, which can be further increased with service-oriented elasticity strategies.
- An elasticity threshold based on client response time and a targeted conservative value can further reduce over-provisioning of resources and maximize the cost savings.

Our next plans include:

- Extension of this study based on larger deployments and more complex applications
- Implementation of a microservices orchestration platform that predicts dynamic load and performs elasticity to maximize infrastructure cost savings for a given service performance target
- Experimentation in novel large-scale test-beds

Thank You