Testing Microservices Architecture-Based Applications: A Systematic Mapping Study Protocol

Table of Contents

Testing Microservices Architecture-Based Applications: A Systematic Mapping Study Protocol

1. Motivation

According to SWEBOK, software testing is composed of dynamic verifications that evaluate if a system provides expected behaviors on a finite set of test cases, suitably selected from the usually infinite execution domain. In distributed systems, testing becomes a complex task. Distributed systems serve a specific requirement, so its architecture is built in a way to serve that business need.

The number of services, inter-communication processes, dependencies, instances, network communication, and other variables influence the testing methodology. In specific distributed systems contexts, like microservices, new benefits have been introduced in order to provide significant features to microservices-based systems such as the ability to independently deploy, scale and maintain each component and parallelize development across multiple teams. These features, however, have involved reconsidering the testing strategies that applied for distributed systems as well as monolithic systems.

Academia and industry discuss many approaches for managing the testing complexity in microservices-based applications. These approaches cover from multiple independently deployable components as well as how to have tests in order to check if the system remains correct despite having multiple teams, each acting as guardians for different services. Apparently, the industry is inclined by using robust and well-known methods, and academia seeks to find new testing methods.

2. Problem statement

Although both academia and the industry have proposed and discussed testing solutions, it is not clear which testing approaches have been proposed for microservices-based systems explicitly. With the increasing interest in the development of MSA-based applications, it is important to systematically identify, analyze, and classify the publication trends, research themes, approaches, and challenges in the context of testing MSA-based applications.

3. Research goal and questions

The goal of this SMS is to *analyze the peer-reviewed literature concerning publication trends, research themes, approaches, and challenges in the context of testing of MSA-based applications.* To conduct the Systematic Mapping Study (SMS) on the testing of microservices-based applications, we formulated the following Research Questions (RQs) presented in Table 1.

Table 1. Research Questions and their Rationale

#	**Research Question**	**Rationale**
RQ1	What are the existing research themes on testing of microservices-based applications and how they can be classified and mapped?	By answering this RQ, we aim to establish the foundation for systematic analysis of the existing research on testing of microservices through a taxonomy of research themes and sub-themes. In this regard, we applied the approach proposed by Braun et al.[1] to identify the main research themes related to the primary studies.
RQ2	What testing approaches have been proposed and used for MSA-based applications?	This RQ aims to describe techniques, procedures, methods, and other proposals used in the process of evaluation of MSA-based applications.

RQ3	Which testing-related challenges have been reported in the primary studies over the years?	This RQ aims to illustrate the challenges reported by primary studies concerning testing on MSA-bases applications. Furthermore, in this RQ, we discussed the evolution of the challenges over the years.

4. Search process

The search process for this study is divided into two phases. First is the primary search, and the second is snowballing. Both phases are briefly explained in subsequent subsections.

4.1. Primary search

During this phase, we will execute the predefined search string from January 2008 to March 2019 on selected databases. Both search string and selected databases are shown in Table 2. we will select the relevant studies by following the following steps.

Step 1- Studies extraction. During this step, we will execute the search string on selected databases and will store the obtained information from database sheets. The output of this step will be an MS Excel sheet with possible Data (D) items regarding study code (D1), studies title (D2), authors list (D3), study publication year (D4), study venue (D5), study publication type (D6), research type (D7), research themes (D8), testing approaches (D9), and challenges (D10).

Step 2- Screening studies through title: After collecting the relevant information from all databases. According to our previous experience, there is a possibility that one study may be published in multiple venues. For example, one study published on ACM also published on IEEE and EI Compendex. For the identification of such duplicate studies, we will organize the collected studies in ascending/descending order, and after that, we will remove the duplicate titles. During this step, we also removed those studies titles that not matched with our research topic (i.e., MSA testing).

Step 3 – Screening studies through abstract reading: During this step, from avoiding the bias, two researchers will review the pool of studies compiled after Step 2. Each researcher will go through the abstract of every study and will rank the paper as "relevant", "irrelevant," or "doubted". We will also review the shortlisted papers by using the keywording technique recommended by Peterson's [2]. During the keywording technique, both reviewers will read abstracts and look for keywords and concepts that reflect the contribution of the paper. While doing so, the reviewer also identifies the context of the research. When this is done, the set of keywords from different papers are combined to develop a high level of understanding about the nature and contribution of the research. There is a possibility, we may find some papers where both authors cannot reach a consensus, and in such cases, we will refer those to the third researcher for getting his/her opinion.

Step 4 - Applying inclusion and exclusion criteria: We will read the full body text of the studies and will decide either this paper is relevant or not according to predefined screening criteria (see Table 3). After this step, we will have the final set of studies for data extraction to answer our RQs.

4.2. Snowballing search

We will apply both forward and backward snowballing techniques on the final set of selected studies (after step 4) according to the guideline in [3]. In order to identify the suitable studies through snowballing, we need to inspect the citations (i.e., backward snowballing) and the references (i.e., forward snowballing) of each selected study.

4.3. Search string and databases

The search string and selected databases for relevant studies are organized in Table 2

Table 2. Selected databases and search string

Search String

(microservice* OR micro service* OR micro-service* OR microservices architect* OR microservices design) AND test*)		
Databases		
Database	**Links**	**Targeted search area**
ACM Digital Library	http://dl.acm.org/	Paper title, abstract
IEEE Explore	http://ieeexplore.ieee.org/	Paper title, keywords, abstract
Springer Link	http://link.springer.com/	Paper title, abstract
Science Direct	http://www.sciencedirect.com/	Paper title, keywords, abstract
Wiley InterScience	http://onlinelibrary.wiley.com/	Paper title, abstract
EI Compendex	https://www.engineeringvillage.com/	Paper title, abstract
ISI Web of Science	https://login.webofknowledge.com	Paper title, keywords, abstract

4.4. Study screening criteria

We will evaluate the collected studies through studies inclusion and exclusion criteria defined in Table 3.

Table 3. Studies inclusion and exclusion criteria

Selection Criteria	**Inclusion Criteria**	**Exclusion Criteria**
Language	• English	• Non-English
Study Type	• Primary studies • Peer review Journals articles, book chapters, conference papers, workshop and symposiums papers	• Secondary studies • Blog, webpages, videos, white papers, technical reports, non-peer review studies (the so-called grey literature).
Study Focus	• Studies that explicitly discuss the testing for MSA based applications • Studies that could provide the information regarding some of data items from D8, D9, and D10	• Studies that discuss the testing for SOA and monolithic applications. • Studies that do not answer the data items from D8, D9, and D10.
Study Duration	• A study published from 2008 to November 2019	• A study published before 2008 and after November 2019

5. SMS documentation and reporting

5.1. Data extraction form

We will design the data form according to data items shown in Table 4. Data items (D1 to D3) will be used to present the overview of the selected studies. For example, index and title of the study. Data items (D4 to D7) will use to answer the demographics. D8 will answer the RQ1, D9, and D10 will answer the RQ2 and RQ3, respectively.

Table 4. Data Extraction Items

Code	**Data Item**	**Description**	**Relevant RQ**
D1	Index	The ID of the study.	Demographics of selected studies
D2	Title	The title of the study.	
D3	Author(s) list	The full name of the authors.	
D4	Year	Publication year of the study.	
D5	Venue	The name of the publishing venue.	
D6	Publication type	Journal, conference, workshop, book chapter, and technical report.	
D7	Research Type	Case study, survey, experiments, validation research, a solution to the proposal.	
D8	Research theme	Execution of guidelines proposed by Braun et al. [1]	RQ1

D9	Testing approaches	What testing strategies (e.g., unit testing, integration testing, consumer testing) are used to evaluate the MSA based system?	RQ2
D10	Challenges	What challenges are reported in the selected studies related to MSA testing?	RQ3

5.2. Study search results

We executed the search string defined in Table 2 and summarized the result in Table 5. The study search duration is from January 2008 to November 2019.

Table 5. Study Search and Selection Process

Databases	**Search and Studies Selection Steps**			
	Step 1	**Step 2**	**Step 3**	**Step 4**
ACM Digital Library	421	198	1	1
IEEE Explore	477	264	30	19
Springer Link	1139	76	18	7
Science Direct	354	12	5	1
Wiley InterScience	79	10	1	0
EI Compendex	Search results on this database were the same as Science Direct			
ISI Web of Science	11	2	0	0
Results	2481	562	55	**28**
Snowballing				**5**
Total Results				**33**

5.3. Data synthesis

We will use descriptive statistics for analyzing the data generated against the data items D1 to D7. We also expect a large amount of qualitative data through data items D8 to D10 mostly comprise of free text description (e.g., study themes, testing approaches,). Therefore, we have a plan to analyzed the qualitative data through thematic analysis guidelines proposed in [1]. These guidelines are consisting of (i) familiarizing with data, (ii) generating initial codes, (iii) searching for themes, (iv) reviewing themes, and (v) naming themes.

5.4. Result reporting

Once we finish the data extraction and analysis of the data, then we will report the results. The possible results will cover and highlight the gaps for MSA based testing approaches, tools, challenges, and their solutions.

5.5. Schedule

We will try our best to meet the following goals according to the designated time.

Study start date: August 06, 2019

Study duration: 4 months

Study end date: March 11, 2020

Study Schedule

#	**Task(s)**	**Estimated duration**
1	Protocol writing	2 weeks
2	Extracting the papers (first phase)	3 weeks
3	Extracting the paper (second phase)	1 week
4	Extracting the paper (third phase)	1 week

5	Reading the paper and extracting the data	3 weeks
6	Synthesizing the data	2 weeks
7	Writing the study draft	4 weeks

References

[1] V. Braun and V. Clarke. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 2 (2006), 77-101.

[2] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. 2008. Systematic mapping studies in software engineering. Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE) (2008), 68-77.

[3] C. Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE) (2014), Article No. 38.