Can big data systems benefit from microservices patterns ?

Pouya Ataei

July 25, 2022

1 Research components:

1. Systematic literature review (following the guidelines of PRISMA and Barbara An Kitchenham)
2. Improved searched strategy by using PRISMA-S
3. Thematic synthesis using Cruzes, D. S. & Dybå approach
4. Capturing microservices patterns following the TOGAF template
5. Understanding the current state of big data architectures
6. Mapping the microservices patterns against big data architectures and see if they can solve some of its issues
7. Discussion
8. Threat to validity
9. Conclusion and further research

2 Timeline:

- two systematic literature review - 1 month (end of June) 10th June - 10th July
- Thematic synthesis (1 week) - 6th July - 13th July
- Capturing patterns in the synthesis (1 week) - 13th July - 20th July
- Critical discussion (1 week) - 20th July - 27th July
- Polish and final edits (1 week) - 27th July - 3rd August

3 Hard Deadlines:

- IEEE Big Data - Aug 20, 2022
- Journal of big data - Aug 30, 2022

4 Chosen Databases:

- IEEE Explore
- ScienceDirect
- SpringerLink
- ACM library
- MIS Quarterly –¿ replace with JSTOR?
- Elsevier
- Scopus
- Aisel
- Wiley? (maybe as an addition?)

5 SLR on Microservices

5.1 Keywords

- at least microservice (or derivates)has to always be in the title:
- microservice* AND pattern*
- microservice* AND architecture*
- microservice* AND design*
- microservice* AND building block*
- microservice* AND best practice*

6 SLR on Big data architecture (following the same methodology by paper published to ACIS (Pouya Ataei - Alan Litchfield) and extend it for the years 2020-2022)

7 Phase 1: search

7.1 Progress report (date):

we have found n number of paper with the following search strategy, and n number has been deduplicated.

Daniel to write here the abnormalies and challenges faced in deduplication and study filtering in the first phase

8 Phase 2: developing inclusion, exclusion criteria, quality framework

8.1 Progress report (date):

we have found n number of paper with the following search strategy, and n number has been deduplicated.

9 Inclusion and Exclusion criteria:

9.1 Inclusion

- Primary and secondary studies between Jan 1st 2012 and June 19th 2022
- The focus of the study is on microservices patterns, and microsrvices architectural constructs.
- Scholarly publications such as conference proceedings and journal papers

9.2 Exclusion

- Studies that are not written in English
- Informal literature surveys without any clearly defined research questions or research process
- Duplicate reports of the same study (a conference and journal version of the same paper). In such cases, the conference paper was removed.
- Short papers (less than 6 pages)

10 Phases of the selection

- Pooling literature - Done
- Duplicate removal - Done
- Removal based on publication types - Done
- Scanning studies titles based on Inclusion and Exclusion criteria - Done
- Scanning studies abstract and title based on Inclusion and Exclusion criteria - Done
- (sub-phase) - Scanning introduction and conclusion parts of the studies based on Inclusion and Exclusion criteria (this was the phase in which we removed non-English studies, we scanned a bit more if needed)
- Assessing the studies based on the quality framework phase 1 (in phase 1, we run the studies against first three questions for minimum quality threshold for all studies, this is done through a questionnaire, and each study must conform to at least 75% of the criteria, while the inter-rater reliability is above 75%), (in phase 2, we run the studies against the second criteria (Rigour), with the exact same statistical condition as the phase 1) (in phase, we run the studies against the third and fourth criteria, with the exact same statistical condition as the phase 1)

11 Quality Framework:

1. Minimum quality threshold:
   (a) Does the study report empirical research or is it merely a 'lesson learnt' report based on expert opinion ?
   (b) The objectives and aims of the study is clearly communicated, including the reasoning for why the study was undertaken ?
   (c) Does the study provide with adequate information regarding the context in which the research was carried out ?
2. Rigour:
   (a) Is the research design appropriate to address the objectives of the research ?
   (b) Is there any data collection method used and is it appropriate ?
3. Credibility:
   (a) Does the study report findings in a clear and unbiased manner ?

   ‘’
4. Relevance:
   (a) Does the study provides value for practice or research?

12 Papers to use in the future (potentially)

1. Architecting with microservices: A systematic mapping study
2. Microservices Anti-patterns: A Taxonomy

13 Daniel note

While conducting the first steps of our filter process, we encountered several hurdles that shall be highlighted to ensure transparency, especially since they can slightly affect the number of remaining entries after those initial phases. However, the final set of literature was not impacted and, therefore, those factors did not pose a threat to the studies validity.

For once, since not all entries of the combined literature list compiled from all the used sources specified a digital object identifier (DOI), the duplicate removal had to be conducted based on the publication title. Yet, in some rare cases, there were duplicates for which the spelling of the title was slightly altered, and which were, therefore, not detected in the initial duplicate removal phase. Instead, they were only identified during the scanning of the title. Furthermore, in SpringerLink, conference papers are classified as book chapter, since conference proceedings are published as books.

This makes them indistinguishable from real book chapters, when only looking at the metadate. Book chapters are, however, not part of the search's scope. Consequently, the removal of book chapters for SpringerLink could only be processed when inspecting the respective publications. To slightly reduce the effort, it was decided to only do this for those publications that passed the filtering by title. As a result, there are some publications in that phase that, in theory, should have already been removed in the previous step.

14 Patterns Format:

1. Name
2. Description (this includes context and problem (only for conference paper))

15 Pattern Catalog:

1. ~~Self-contained service~~
2. ~~Decomopose by business capability~~
3. Database per service
4. Shared database
5. Event sourcing
6. Multiple service instances per host
7. ~~Serverless deployment~~
8. Containerized
9. API gateway
10. Self registration
11. Service discovery
12. Client-side discovery
13. Server-side discovery
14. Circuit breaker
15. Synchronous communication
16. ~~Publish/subscribe communication~~
17. ~~Combination of HTTP and Message queue~~
18. ~~Communication using message-oriented middleware~~
19. ~~Asynchronous communication~~
20. ~~Point-to-point communication~~
21. ~~Communication using binary protocols~~
22. Bulkhead pattern
23. Local database proxy
24. Local sharding-based router
25. Command and query responsibility segregation

26. ~~Scalable store~~
27. ~~Secure channel~~
28. ~~Change code dependency to service call/adapter microservice~~
29. Service registry
30. Service registry client
31. Service locator
32. Event notification
33. ~~REST integration~~
34. Competing consumers
35. Pipes and filters
36. ~~Asynchronous messaging~~
37. Request-reaction
38. Strangler
39. Anti-corruption layer
40. Externalized configuration
41. Self-containment of services
42. ~~Container~~
43. ~~Deploy cluster and orchestrate containers~~
44. ~~Microservices Devops~~
45. Database is the service
46. Enable continuous integration
47. Self-contained systems
48. Bulkhead
49. Circuit breaker
50. Priority queue
51. ~~Results cache~~
52. ~~Page cache~~
53. ~~Key-value store~~
54. ~~Correlation ID~~
55. Log aggregator

56. ~~Loading balancer / load-balancing~~
57. Ambassador
58. Sidecar
59. Gateway aggregate
60. Gateway offloading
61. ~~Asynchronous query~~
62. ~~Asynchronous completion token~~
63. ~~Edge server~~
64. ~~Internal load balancer~~
65. ~~External load balancer~~
66. ~~Health check~~
67. ~~Monitor~~
68. ~~Consumer-driven contracts~~
69. Tolerant reader
70. Aggregator
71. Backend for Frontend
72. API Composition
73. Microservice Chassis/Service Host
74. ~~Service-per-container~~
75. ~~Service deployment platform~~
76. ~~Serverless deployment~~
77. Saga transaction management
78. Gateway routing
79. External configuration store
80. Static content hosting
81. Computer resource consolidation
82. Leader election
83. Database cluster
84. Hybrid

16 Paper Sections

1. Introduction (1/5 to 2 columns) –¿ 0.75 page
2. Background (1 columns) –¿ 0.5 page
3. Research Methodology (2/3 4/5 columns) –¿ 3 pages
4. Results (6 columns) –¿ 3 pages
5. Threats to validity (half a column)
6. Discussion (1 column)
7. Conclusion (1 column)

17 Plan

17.1 Pouya

To research about requirements engineering and potentially a formal method of describing big data requirements

17.2 Daniel

To start writing the research methodology chapter