

The state of big data reference architectures: a systematic literature review

Abstract

This template helps you to create a properly formatted L^AT_EX manuscript.

Keywords: `elsarticle.cls`, L^AT_EX, Elsevier, template

2010 MSC: 00-01, 99-00 s

1. Introduction

The rapid development of software technologies, the proliferation of digital devices and networking infrastructure of today, have by and large, augmented user's capability to generate data [1]. In the age of information, users are
5 unceasing generators of structured, semi-structured, and unstructured data that if collected and crunched correctly, may reveal game-changing patterns [2].

The unprecedented proliferation of data have emerged a new ecosystem of technologies; one of these ecosystems is big data [3]. Big data is a term emerged to describe large amount of data that comes in various forms from different chan-
10 nels. Within the years, big data has attained a lot of attention from academia and industry, and many strive to benefit from this new material. Howbeit, adopting big data requires the absorption of great deal of complexity and many traditional systems cannot cope with characteristics of this domain.

A recent survey published by Databricks in partnership with MIT Technol-
15 ogy Review Insights, stated that only 13% of companies excel at delivering on their data strategy [4]. In the same vein, Vintage Partners highlighted that only 24% of companies have successfully adopted big data [5]. Sigma computing report presented that 1 in 4 business experts have given up on getting insights they needed because the data processing took too long [6]. Moreover, Gartner

20 approximated that only 20% of companies have successfully adopted big data.

Some of the most highlighted challenges of big data is 'lack of business context', 'organizational challenges', 'big data architecture', 'data engineering', 'rapid technology change', and 'lack of talent' [7]. Whereas similar issues may exist in other domains, it is exacerbated when it comes to big data systems. This
25 is due the inherent complexity of big data engineering, the need for real-time processing, the scalability requirement of these systems, and the sensitivities around data.

Today, majority of big data systems are designed underlying ad-hoc and complicated architectural solutions [8], that do not seem to adhere to similar
30 patterns. This will challenge software architects to design a suitable solution for any given context, creates a foundation for an immature architectural decision, and does not promote the growth and development of BD systems as a whole.

Therefore, since the approach of ad-hoc design to big data systems is undesirable and leaves many engineers in the dark, there is a need for more software
35 engineering research for big data systems. To this end, this study presents a systematic literature review (SLR) on big data (BD) reference architectures (RAs).

2. Why reference architectures?

Conceptualization of the system as a reference architecture, helps with un-
40 derstanding of the system's key components, behavior, composition and evolution of it, which in turn affect quality attributes such as maintainability, scalability and performance [9]. Therefore RAs can be a good standardization artefact and a communication medium that not only results in concrete architectures for big data systems, but also provide stakeholders with unified elements and
45 symbols to discuss and progress big data projects

This approach to system development is not new to practitioners of complex system. In software product line (SPL) development, RAs are utilized as generic artifacts that are instantiated and configured for a particular domain of systems

[10]. In software engineering, IT giants like IBM have referred to RAs as the
50 'best of best practices' to address complex and unique system design challenges
[9]. In other international standardization, RAs have been repeatedly used to
standardize an emerging domain, a good example of this is BS ISO/IEC 18384-1
RA for service oriented architectures [?].

- document style
- 55 • baselineskip
- front matter
- keywords and MSC codes
- theorems, definitions and proofs
- lables of enumerations
- 60 • citation style and labeling.

3. Front matter

The author names and affiliations could be formatted in two ways:

- (1) Group the authors per affiliation.
- (2) Use footnotes to indicate the affiliations.

65 See the front matter of this document for examples. You are recommended to
conform your choice to the journal you are submitting to.

4. Bibliography styles

There are various bibliography styles available. You can select the style of
your choice in the preamble of this document. These styles are Elsevier styles
70 based on standard styles like Harvard and Vancouver. Please use BibT_EX to
generate your bibliography and include DOIs whenever available.

Here are two sample references: [? ?].

5. Improvements

1. The current writing style looks like a summary description, lacks new
75 insight on the topic. The overall contribution needs to be enhanced.
2. The author raises seven research questions, but how does the author develop these questions? Are these real questions that have never been discussed? The research questions need to be developed according to the literature. In this way, we can realize what is the gap on this topic.
- 80 3. The inclusion criteria and exclusion criteria are ambiguous and questionable. Is it possible for readers to reproduce this study according to these criteria? Did the author perform the reliability and validity tests? The author needs to provide more detail about the review methodology.
4. In general, each larger-scale system requires a more understanding of architectural components, owing largely to the complex nature of system
85 architects. However, I cannot find a case that the authors demonstrate the uniqueness of BD systems, and the actual development challenges in BD systems.
5. Although this study adapts SLR approach, it should not completely miss
90 a literature reviewing section. It is necessary to provide to the reader the preliminary details which are necessary to understand the purpose of this study, techniques and key concerns of the various research work that the authors have reviewed. There is no theoretical argument to support the development of the research questions.
- 95 6. The findings yielded by investigating the research questions of this SLR should constitute many discussion points around the research and practice of BD systems. However, the manuscript is completely missing a discussion section. One should expect that the results of SLR can inform the current knowledge and provide several research directions for future
100 research.
7. Last, one of the core challenges with the paper is to situate it within an ongoing scholarly conversation. The authors currently reference a fairly

diverse set of papers, but remain at a fairly abstract level when it comes to elaborating how your work builds upon and expands existing work. In turn, this makes it difficult to appreciate theoretical implications of your work.

References

References

- [1] B. Bashari Rad, N. Akbarzadeh, P. Ataei, Y. Khakbiz, Security and privacy challenges in big data era, International Journal of Control Theory and Applications 9 (43) (2016) 437–448.
- [2] P. Ataei, A. T. Litchfield, Big data reference architectures, a systematic literature review (2020).
- [3] B. B. Rada, P. Ataeib, Y. Khakbizc, N. Akbarzadehd, The hype of emerging technologies: Big data as a service (2017).
- [4] Databricks.
URL <https://databricks.com/>
- [5] N. Partners, Big data and ai executive survey 2021 (2021).
URL https://www.supplychain247.com/paper/bi_data_and_ai_executive_survey_2021/pragmadik
- [6] S. Computing, Bridging the gap between data and business teams (2020).
URL <https://www.sigmacomputing.com/resources/data-language-barrier/>
- [7] B. B. Rad, P. Ataei, The big data ecosystem and its environs, International Journal of Computer Science and Network Security (IJCSNS) 17 (3) (2017) 38.
- [8] I. Gorton, J. Klein, Distribution, data, deployment, STC 2015 (2015) 78.

- [9] R. Cloutier, G. Muller, D. Verma, R. Nilchiani, E. Hole, M. Bone, The concept of reference architectures, *Systems Engineering* 13 (1) (2010) 14–27.
- [10] M. Derras, L. Deruelle, J.-M. Douin, N. Levy, F. Losavio, Y. Pollet, V. Reiner, Reference architecture design: A practical approach, in: *IC-SOFT*, pp. 633–640.