

Draft chapter

Literature Review

Fred Spoons

5 July 2020

# Chapter 2

## Literature Review

### 2.1 Introduction

The literature review design of this PHD research constitutes of 3 parts, 2 systematic literature reviews (SLR) on the topics ‘Big data reference architectures’ and ‘E-commerce systems’ and one generic literature review on the topic ‘Big data’.

Systematic literature reviews take shape by embarking on an extensive search for topic-related articles within the years 2010-2020. Most literature chosen for the purposes of this research are within the years 2016-2020 as they provided with recent, and more relevant information. Albeit, some old studies dating back to 2010, helped clarifying some basic matters that existed and how they correlated to big data. world most renowned online libraries for quality research have been selected such as IEEE, MIS Quarterly, Science Direct, Elsevier, Springer, ACM, AISeL and Emerald insight.

Every library provided with a vast sea of research and inordinate amount of information to absorb. Arguably, different publications provided with different sort of mental framework and so did the authors. For instance, it’s been found that many high-quality information system researches are published in MIS quarterly, whereas Elsevir and SpringerLink provided with quality big-data literature.

A combination of long-tail and short-tail keywords are chosen to target literature that are related to the current state of art. Keywords chosen for 'Big data reference architectures' SLR are 'big data reference architectures', and 'reference architectures'. Keywords chosen for 'E-commerce systems' are 'e-commerce system architectures', 'smart e-commerce systems', and 'e-commerce and big data'. Each systematic literature review is conducted in a span of three weeks.

In what follows, first the generic big data literature review will be conducted, second 'big data reference architecture' SLR and finally 'E-commerce systems' SLR will take place.

## 2.2 State of the art

We've come a long way with technology, and specifically software development. In fact, the rapid advancements left many spaced out. From the emergence of the first computer Eniac in 1946 to 8-core 5.0GHz processing core speed in 2019; From document-oriented waterfalls to agile two-weeks sprints; from punch cards to fancy transpilers and dynamic programming languages. Computers were first perceived as calculation engines and has been used to focus entirely on algorithms and mathematics. It was during the mid-1950, that it became commercially available and businessmen start to pick it up to produce value for business. Along the lines, once people started using computers for real-life purposes, many leftover data has been produced, as these data increased, people started realizing the value of it and began to store it (?, ?).

That's where the industry came up with a concept of a Database Management System (DBMS), and humanity began to store data for various purposes. In 1968, as a result of a NATO-sponsored conference, the term software engineering emerged, referring to a highly systematic approach to software development and maintenance (?, ?).

Since the beginning of 1968, the advancement began on the areas of tools generation, testing, automation and systematizing. During the same years, in 1960s the history of computer hardware started by conversion of vacuum tubes to solid-states. Today, the word 'bug' is quite a common phrase among engineers and programmers to refer to a fault, failure or a flaw. We owe this word to a literal moth that was caught inside a tube before the transition to solid states. It is hardly conceivable that we've progressed from absolutely no understanding for data to devices that can produce zettabytes of them in a span of 60 years. Along this track, software engineering has passed several major phases. Recent polyglot approaches with nascent lambda functions, functional paradigms and micro-services have come to take the industry by storm. This is the only time in human history, where the computing resources and the necessary data is available to harness the hidden patterns behind every momentum or dynamism. Being so focused on development of more maintainable and scalable software, and microchips and hardware's and devices that can perform faster and last longer, we have lost the track of the output of all these entities and peripherals, and that's the void that current industry is facing. Abundance of computer power, the emergence of open source community, and the ubiquity of internet has brought us with a new material to harness. A material, that is complex and random in nature.

It was not until 2005 that the term big data has been coined (?), and Web 2.0 emerged which referred to a large set of data that is impossible to process with the traditional data management systems. Within the same year, Yahoo created Hadoop, Google came up with MapReduce. In 2009, the Indian government took a revolutionary step and decided to take an iris scan of its 1.2 billion inhabitants. In 2011 McKinsey published the title "Big Data: the next frontier of innovation" and startups and companies started investing heavily in this field. The big data revolution is ahead of us, and yet there is a big chasm both in practice and academia (?).

## 2.3 Big data

### 2.3.1 What is big data?

To define big data for the course of this PHD thesis, we will first look at available definitions in academia.

? define big data as “the amount of data which is beyond technology’s capability to store, manage and process efficiently.” referred to big data as “the use of large data sets to handle the collection or reporting of data that serves business or other recipients in decision making”.

? define big data as “a term for massive data sets having large, more varied and complex structure with the difficulties of storing, analyzing and visualizing for further processes or results”. Inspired by these definitions, we define big data as “an endeavor to harness the patterns behind vast amount of data for the purposes of improvement, control, and prediction of business matters”.

## 2.4 The Hype of Emerging Technologies

The term big data, was initially coined to refer to the gradual growth and availability of data (? , ?).

The ubiquity of digital devices and capability of users to produce different forms of data, have consolidated the interconnected links among suppliers, customers, affiliates, partners, and stakeholders (? , ?). With recent emergence of 5G technology and its launch

---

<sup>1</sup>A transpiler is a sort of a compiler that translates source codes from one language to another, or another version of the same language. For example Babel (a Javascript Library ) transpiles the latest syntax of Javascript ( ES6 ) into older version of it ( ES5 ), thus all the browsers can support the system.

in the UK, we are experiencing a fundamental network shift that is unprecedented in human history (?, ?).

Opposed to general belief of 5G being only faster than its elder brother 4G, 5G has come to offer bi-directional large bandwidth shaping, large broadcasting of data in gigabits which supports wearable devices with AI capabilities, pervasive networks providing ubiquitous computing (the user can be seamlessly connected to several wireless access technologies), traffic statistics, IPV6 utilization and finally 25Mbps of connectivity speed (?, ?).

In a world where we have the average processing power of 1.5 GHz on smart phones and up to 8 GHz on desktops running on network infrastructures that will support up to 25Mbps of transmission per second, data becomes the new oil, the atom, the dot that lays the foundation of the nexus (?, ?).

# References

- Bughin, J. (2016). Big data, big bang? [Journal Article]. *Journal of Big Data*, 3(1), 2. doi: 10.1186/s40537-015-0014-3
- Gohil, A., Modi, H. & Patel, S. K. (2013). 5g technology of mobile communication: A survey [Conference Proceedings]. In *2013 international conference on intelligent systems and signal processing (issp)* (p. 288-292). IEEE. doi: 10.1109/issp.2013.6526920
- Grad, B. & Bergin, T. J. (2009). Guest editors' introduction: History of database management systems [Journal Article]. *IEEE Annals of the History of Computing*, 31(4), 3-5. doi: 10.1109/mahc.2009.99
- Holma, H., Toskala, A. & Nakamura, T. (2020). *5g technology: 3gpp new radio* [Book]. John Wiley Sons. Retrieved from [https://www.ebook.de/de/product/37929190/harri\\_holma\\_dr\\_antti\\_toskala\\_takehiro\\_nakamura\\_5g\\_technology.html](https://www.ebook.de/de/product/37929190/harri_holma_dr_antti_toskala_takehiro_nakamura_5g_technology.html)
- Kaisler, S., Armour, F., Espinosa, J. A. & Money, W. (2013). Big data: Issues and challenges moving forward [Conference Proceedings]. In *2013 46th hawaii international conference on system sciences* (p. 995-1004). IEEE. doi: 10.1109/hicss.2013.645
- Long, C. (2015). Data science and big data analytics: Discovering, analyzing, visualizing and presenting data [Journal Article]. *Indianapolis, Indiana*. doi: 10.1109/bgdds.2018.8626811
- Lycett, M. (2013). *'datafication': making sense of (big) data in a complex world* (Vol. 22). Taylor & Francis. doi: 10.1057/ejis.2013.10
- McKinsey, G. et al. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, 158-184. doi: 10.7591/9781501734328-007
- Rad, B. B. & Ataei, P. (2017). The big data ecosystem and its environs [Journal Article]. *International Journal of Computer Science and Network Security (IJCSNS)*, 17(3), 38. doi: 10.1007/978-3-319-94301-5\_16
- Sagiroglu, S. & Sinanc, D. (2013). Big data: A review [Conference Proceedings]. In *2013 international conference on collaboration technologies and systems (cts)* (p. 42-47). IEEE. doi: 10.1109/cts.2013.6567202
- Srivastava, R. (2018). Big data: Issues and challenges [Journal Article]. *International Journal Of Scientific And Innovative Research*(6), 1. doi: 10.1007/978-981-13-8759-3\_10

- 
- Wirth, N. (2008). A brief history of software engineering [Journal Article]. *IEEE Annals of the History of Computing*, 30(3), 32-39. doi: 10.1109/mahc.2008.33